Statistical Theory

James Marshall Reber

May 3, 2017

Contents

Remark. *Unlike the Probability notes, these notes to just aim to be as concise as possible so that they can be an effective reference on the test. I plan on including all of the definitions and theorems from the book, as well as anything important from the class lectures and the homework solutions/questions. If there's time, I'll include as many examples as I can. You probably should use these in tangent with the class notes.*

Chapter 1

Reference Tables

Remark. *If you plan on using the t-table, I've omitted some values. I don't think this will be an issue, but if this bugs you, you should find another table to use on the internet.*

Name	Density	Domain	Expected Value	Variance	Parameters	When Used
Uniform	$1/(\theta)$	$0 \leq x \leq \theta$	$(\theta)/2$	$(\theta)^2/12$	θ	Over intervals
Exponential(λ)	$\lambda e^{-\lambda x}$	$x \geq 0$	$1/\lambda$	$1/\lambda^2$	λ is average number of successes	Wait time until 1st event
Gamma(α, β)	$\frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}$	$x \geq 0$	α/β	α/β^2	β is average number of successes, α is number of things	Wait time until α-th event
Beta(α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 \leq x \leq 1$	$\alpha/(\alpha+\beta)$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	Usually given constraints	Bayesian statistics
Normal	$\frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$	$-\infty < x < \infty$	μ	σ^2	μ =expected val, σ^2 = variance	Central Limit theorem and applic.
χ^2-distribution	$\frac{1}{2^{v/2}\Gamma(v/2)} x^{\frac{v}{2}-1} e^{-\frac{x}{2}}$	$x > 0$	v	$2v$	$v > 0$	You use this only really for tricks in this course
Pareto	$\alpha(1+x)^{-(\alpha+1)}$	$\alpha > 0$	$\frac{1}{\alpha-1}$	$\frac{2}{\alpha^2-3\alpha+2}$	$\alpha > 0$.	It's never been stated what the use is.

Table 1.1: Named Continuous Random Variables

Note: $\Gamma(r) = (r-1)!$.

Name	Mass	Expected Value	Variance	When Used
Bernoulli	$p_X(1) = p$, $p_X(0) = q$	p	pq	If there is 1 success or failure.
Binomial	$\binom{n}{x} p^x q^{n-x}$	np	npq	If we're measuring the amount of successes in n trials.
Geometric	$q^{x-1}p$	$1/p$	q/p^2	Measuring the amount of trials until the first success
Negative Binomial	$\binom{x-1}{r-1} q^{x-r} p^r$	r/p	qr/p^2	Measuring the amount of trials until the r-th success.
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	Measuring the number of events in a period.
Hypergeometric	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$n\frac{M}{N}$	$n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$	Measuring the number of good things selected.
Discrete Uniform	$\frac{1}{N}$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	If everything is equally likely.

Table 1.2: Named Discrete Random Variables

Table 1.3: Counting equations

	Sampling with replacement	Sampling without replacement
Order matters	n^r	$\frac{n!}{(n-r)!}$
Order does not matter	$\binom{n+r-1}{r}$	$\binom{n}{r}$

Table 1.4: Random Variable Facts

	Discrete	Continuous
Probability Function	Mass (probability mass function; PMF)	Density (probability density function; PDF)
	$0 \leq p_X(x) \leq 1$	$0 \leq f_X(x)$
	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)dx = 1$
	$P(0 \leq X \leq 1) = P(X = 0) + P(X = 1)$ if X is integer valued	$P(0 \leq X \leq 1) = \int_0^1 f_X(x)dx$
	$P(X \leq 3) \neq P(X < 3)$ when $P(X = 3) \neq 0$	$P(X \leq 3) = P(X < 3)$
cumulative distribution function (CDF)	$F_X(a) = P(X \leq a) = \sum_{x \leq a} P(X = a)$	$F_x(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx$
named distributions	Bernoulli, Binomial, Geometric, Negative, Binomial, Poisson, Hypergeometric, Discrete Uniform	Continuous Uniform, Exponential, Gamma, Beta, Normal
expected value	$\mathbb{E}(X) = \sum_x x p_X(x)$, $\mathbb{E}(g(X)) = \sum_x g(x) p_X(x)$	$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx$, $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$
variance	$\text{Var(X)} = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$	$\text{Var(X)} = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
covariance	$\text{Cov(X,Y)} = E(XY) - E(X)E(Y)$	$\text{Cov(X,Y)} = E(XY) - E(X)E(Y)$

Table 1.5: Continuity Correction Table

Strictly less than	Subtract 0.5
Less than or equal to	Add 0.5
Greater than or equal to	Subtract 0.5
Strictly greater than	Add 0.5

Table 1.6: Type I and Type II error

	Actual Truth	
Conclusion	True	False
True	Correct	Type II error
False	Type I error	Correct

Table 1.7: Z-table

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

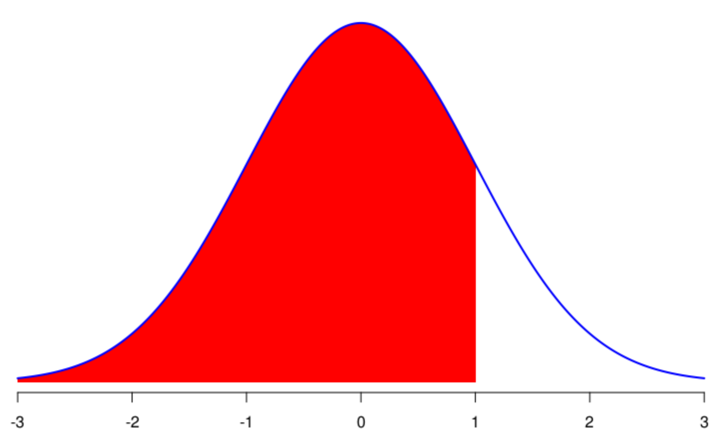

Table 1.8: t-table

df	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

Table 1.9: Chi-Squared Percentage Points

ν	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

Table 1.10: Chi-Squared Percentage Points

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

Chapter 2

Review

Variance, Covariance, and Correlation

Definition 2.0.1. (Variance) The variance of a random variable X is the quantity

$$\sigma_x^2 = \mathrm{Var}(X) = E\big((X - \mu_X)^2\big),$$

where $\mu_X = E(X)$ is the mean of X. Alternatively, we have

$$E(X^2) - E(X).$$

Theorem 1. *Let X be any random variable with expected value $\mu_X = E(X)$ and variance $Var(X)$. Then the following hold true:*
(a) $Var(X) \geq 0$.
(b) If a and b are real numbers, $Var(aX + b) = a^2 Var(X)$.
(c) $Var(X) \leq E(X^2)$.

Definition 2.0.2. (Covariance) The covariance of two random variables X and Y is given by

$$Cov(X, Y) = E\big((X - \mu_X)(Y - \mu_Y)\big),$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

Theorem 2. *Let X, Y, and Z be three random variables. Let a and b be real numbers. Then*

$$Cov(aX + bY, Z) = a\,Cov(X, Z) + b\,Cov(Y, Z).$$

Note there is also a symmetry; i.e.

$$Cov(Z, aX + bY) = a\,Cov(Z, X) + b\,Cov(Z, Y).$$

Theorem 3. *Let X and Y be two random variables. Then*

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Corollary 3.1. *If X and Y are independent, then $Cov(X, Y) = 0$.*

Theorem 4. *(a) For any random variables X and Y,*

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

(c) More generally, we have

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X, Y).$$

(b) Even more generally, for any random variables $X_1, \ldots, X_n$,

$$Var\Big(\sum_i X_i\Big) = \sum_i Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j).$$

Corollary 4.1. *(a) If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$.*
(b) If $X_1, \ldots, X_n$ are independent, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$.

Definition 2.0.3. The correlation of two random variables X and Y is given by

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Joint and Conditional Probability

Definition 2.0.4. (Joint Probability Function) Let X and Y be discrete random variables. Then their joint probability function $\rho_{X,Y}$, is a function from $\mathbb{R}^2$ to $\mathbb{R}^1$, defined by

$$\rho_{X,Y}(x,y) = P(X = x, Y = y).$$

Theorem 5. *Let X and Y be two discrete random variables, with joint probability function $\rho_{X,Y}$. Then the probability function ρ_X of X can be computed as*

$$\rho_X = \sum_y \rho_{X,Y}(x,y).$$

Similarly, for the continuous case, we have

$$\rho_X = \int_y \rho_{X,Y}(x,y).$$

Definition 2.0.5. Suppose X and Y are two discrete random variables. Then the conditional probability function of Y, given X, is the function $\rho_{Y|X}$ defined by

$$\rho_{Y|X}(y|x) = \frac{\rho_{X,Y}(x,y)}{\sum_z \rho_{X,Y}(x,z)} = \frac{\rho_{X,Y}(x,y)}{\rho_X(x)}$$

defined for all $y \in \mathbb{R}$ and all x with $\rho_X(x) > 0$. Likewise, for the continuous case, we have

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},$$

which is valied for all $y \in \mathbb{R}$ and for all x such that $f_X(x) > 0$.

Convergence

While the notes say a lot, he said in lecture that we really only care about one kind of convergence.

Definition 2.0.6. (Convergence in Probability) Let $\{X_n\}$ be a sequence of random variables defined on a sample space, Ω. We say that $\{X_n\}$ is convergent in probability to a random variable X defined on Ω iff

$$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0.$$

for any $\epsilon > 0$.

Intuitively, if our sequence of random variables converge in probability, then the probability that our random variable is far away from the value it converges to should be 0. This does not mean that the values are not "far away" from our target value, but rather this should happen essentially never.
An easy way to prove convergence in probability is using Chebychev's Inequality.

Theorem 6. *(Chebychev's Inequality) Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k > 0$,*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Remark. *We equivalently have*

$$P(|X-\mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

and

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

What exactly does Chebychev's inequality say? It guarantees that, for a wide class of probability distributions, "nearly all" values are close to the mean. In other words, a minimum of just 75% of values must lie within two standard deviations and 89% within three standard deviations (in contrast to the 68-95-99.7 rule).

Chapter 3

Likelihood Inference

Remark. *We have something which is called the Likelihood Principle. It says that if two model and data combinations yield equivalent likelihood functions, then inferences about the unknown parameter must be the same.*

Definition 3.0.1. Sufficient Statistic A function T defined on the sample space S is called a sufficient statistic for the model if, whenever $T(s_1) = T(s_2)$, then

$$L(\cdot|s_1) = c(s_1, s_2)L(\cdot|s_2)$$

for some constant $c(s_1, s_2) > 0$.

In other words, it is a sufficient statistic if the likelihood functions are proportional by a constant.

Theorem 7. *(Factorization Theorem) If the density (or probability function) for a model factors as $f_\theta(s) = h(s)g_\theta(T(s))$, where g_θ and h are nonnegative, then T is a sufficient statistic.*

This has a lot of language behind it, but you should think of this less of something super rigorous and more as just whether or not you can separate your data function from your parameter function.

Definition 3.0.2. (Minimal sufficient statistic) A sufficient statistic T for a model is a minimal sufficient statistic, whenever the value $T(s)$ can be calculated once we know the likelihood function $L(\cdot|s)$.

Definition 3.0.3. (Maximum likelihood estimate) We call $\hat{\theta} : S \to \Omega$ satisfying $L(\hat{\theta}(s)|s) \geq L(\theta|s)$ for every $\theta \in \Omega$ is a maximum likelihood estimator, and the value $\hat{\theta}(s)$ is called a MLE (maximum likelihood estimate).

Theorem 8. *If $\hat{\theta}(s)$ is an MLE for the original parameterization and, if Ψ is a 1-1 (injective) function defined on Ω, then $\hat{\psi}(s) = \Psi(\hat{\theta}(s))$ is an MLE in the new parameterization.*

Definition 3.0.4. (Log-likelihood function) For likelihood function $L(\cdot|s)$, the log-likelihood (or log-likelihood) function $l(\cdot|s)$ defined on Ω is given by $l(\cdot|s) = log(L(\cdot|s))$.

Theorem 9. *(Per Dr. Zhang) We will* ***always*** *have that $\frac{\partial}{\partial\theta}l(\theta|s)$ is the MLE, provided it exists.*

Remark. *If you're not in this class, then you should check to make sure that this is a maximum by taking the second derivative.*

Definition 3.0.5. (Mean-squared error) The mean-squared error (MSE for short) of the estimator T of $\psi(\theta) \in \mathbb{R}$ is given by $MSE_\theta(T) = E_\theta\big((T - \psi(\theta))^2\big)$ for each $\theta \in \Omega$.

Theorem 10. *If $\psi(\theta) \in \mathbb{R}$ and T is a real-valued function defined on S such that $E_\theta(T)$ exists, then*

$$MSE_\theta(T) = Var_\theta(T) + \big(E_\theta(T) - \psi(\theta)\big)^2.$$

Definition 3.0.6. (Bias) The bias in the estimator T of $\psi(\theta)$ is given by $E_\theta(T) - \psi(\theta)$ whenver $E_\theta(T)$ exists. When the bias in an estimator T is 0 for every θ, we call T an unbiased estimator of ψ, i.e., T is unbiased whenever $E_\theta(T) = \psi(\theta)$ for every $\theta \in \Omega$.

Definition 3.0.7. (Consistent)A sequence of estimates $T_1, T_2, \ldots$ is said to be consistent (in probability) for $\psi(\theta)$ if $T_n \to \psi(\theta)$ in probability as $n \to \infty$ for every $\theta \in \Omega$.

Here, we see the importance of convergence in probability.

Definition 3.0.8. (Confidence Interval) An interval $C(s) = (l(s), u(s))$ is a γ-confidence interval for $\psi(\theta)$ is $P_\theta(\psi(\theta) \in C(s)) = P_\theta(l(s) \leq \psi(\theta) \leq u(s)) \geq \gamma$ for every $\theta \in \Omega$. We refer to γ as the confidence level of the interval.

This is just a lot of language to describe the standard confidence interval that we're used to. In essence, you should focus on that if we know the variance then the confidence interval is

$$\left[\bar{x} - z_{(1+\gamma)/2}\sqrt{\frac{\sigma^2}{n}}, \bar{x} + z_{(1+\gamma)/2}\sqrt{\frac{\sigma^2}{n}}\right]$$

where γ is our confidence level and

$$z_{(1+\gamma)/2} = \Psi^{-1}\left(\frac{1+\gamma}{2}\right),$$

with Ψ as the CDF of $N(0,1)$. See the table in the first chapter for actual numbers.

Remark. *Note that we do not necessarily require our data to be normal to do this. For example, if we had $(x_1, \ldots, x_n)$ as a sample from a Bernoulli(θ) distribution where $\theta \in [0,1]$ is unknown and we want a γ-confidence interval for θ, then we calculate it using the following formula*

$$\left[\bar{x} - z_{(1+\gamma)/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{(1+\gamma)/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

since, by the CLT, we have that it converges to a normal distribution as we let $n \to \infty$.

If we don't know our variance, then we preform a t-test. Here, the formula is

$$\left[\bar{x} - t_{(1+\gamma)/2,n-1}\sqrt{\frac{s^2}{n}}, \bar{x} + t_{(1+\gamma)/2,n-1}\sqrt{\frac{s^2}{n}}\right]$$

where $t_{(1+\gamma)/2,n-1}$ is the quantile values of the t_{n-1} distribution.

Definition 3.0.9. (Margin of Error) Note that in the equations we have that they are symmetric about $\bar{x}$. We define the margin of error to be the half-length; in other words,

$$z_{(1+\gamma)/2}\sqrt{\frac{\sigma_0^2}{n}}$$

is the margin of error if we know the variance. If we don't know the variance,

$$t_{(1+\gamma)/2,n-1}\sqrt{\frac{s^2}{n}}.$$

is the margin of error.

Remark. *He doesn't seem too concerned with us knowing P-values in relation to hypotheses testing, but I'll include it regardless.*

Two-Sided Hypotheses Testing

If we know the variance, we have that the P-value is

$$P = 2\left[1 - \Phi\left(\left|\frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2/n}}\right|\right)\right].$$

If the P-value is small, then we have evidence that $\bar{x}$ is a suprising value, since this tells us that $\bar{x}$ is out in a tail of our Normal distribution. Hence, we would have evidence to reject the null-hypothesis if this P-value is small; otherwise, we do not have evidence to reject the null hypothesis (make sure to use this wording).

Remark. *We can do this once again with a Bernoulli sample, using the same idea as before. The formula is*

$$P = 2\left[1 - \Phi\left(\left|\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right|\right)\right].$$

If we do not know the variance, we then have that the P-value is

$$P = 2\left[1 - G\left(\left|\frac{\bar{x} - \mu_0}{\sqrt{s^2/n}}\right|; n-1\right)\right],$$

where $G(\cdot; n-1)$ denotes the distribution of the $t(n-1)$ distribution.

One-Sided Hypothesis Testing

We have two cases to consider: if $H_0 : \psi(\theta) \leq \psi_0$ or $H_0 : \psi(\theta) \geq \psi_0$.
If we know the variance and $H_0 : \psi(\theta) \leq \psi_0$, we then have that our P-value is

$$P = 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2/n}}\right).$$

The one-sided confidence interval is then

$$\left[\bar{x} + z_\gamma\sqrt{\sigma_0^2/n}, \infty\right).$$

If we know the variance and $H_0 : \psi(\theta) \geq \psi_0$, we then have that our P-value is

$$P = \Phi\left(\frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2/n}}\right).$$

The one-sided confidence interval is then

$$\left(-\infty, \bar{x} - z_\gamma\sqrt{\sigma_0^2/n}\right].$$

We never discussed the case where we do not know the variance, but presumably it follows very similarly.

Type I error, Type II error, Significance Level, and Power

Remark. *He's mentioned in class that power isn't that important.*

Note that Type I error is simply when we conclude H_1 when in fact H_0 is true, and Type II error is when we conclude H_0 when actually H_1 is true. This table shows their relationship.

Table 3.1: Type I and Type II error

	Actual Truth	
Conclusion	True	False
True	Correct	Type II error
False	Type I error	Correct

Two-Sided Hypotheses

Assume we have $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$, where a is a value to be determined (i.e., it's our confidence level). Then we have the following formulae:
Rejection Region: $C = \{|\frac{\bar{x}-\mu_0}{\sqrt{\sigma_0^2/n}}| > a\}$.
Type I error: P(Conclude $H_1|H_0) = 2\Phi(-a)$.
Type II error: P(Conclude $H_0|H_1$) =

$$\Phi\Big(a + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0}\Big) - \Phi\Big(a - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0}\Big)$$

.
Significance Level: $2\Phi(-a)$.
Power Function: P(Conclude H_1) $= 1 - \Big(\Phi\Big(a + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma_0}\Big) - \Phi\Big(a - \frac{\sqrt{n}(\mu_0-\mu)}{\sigma_0}\Big)\Big)$

One-Sided Hypotheses

First, let's assume we have $H_0 : \mu \leq \mu_0$ and $H_1 : \mu > \mu_0$, where a is a value to be determined. Then we have the following formulae:
Rejection Region: $C = \{\bar{x} \geq a\}$.
Type I error: P(Conclude $H_1|H_0) = 1 - \Phi\big(\frac{a-\mu}{\sqrt{\sigma_0^2/n}}\big)$.
Type II error: P(Conclude $H_0|H_1) = \Phi\big(\frac{a-\mu}{\sqrt{\sigma_0^2/n}}\big)$.
Power Function: P(Conclude H_1) $= 1 - \Phi\big(\frac{a-\mu}{\sqrt{\sigma_0^2/n}}\big)$.
Next, let's assume we have $H_0 : \mu \geq \mu_0$ and $H_1 : \mu < \mu_0$, where a is a value to be determined. Then we have the following formulae:
Rejection Region: $C = \{\bar{x} \leq a\}$.
Type I error: P(Conclude $H_1|H_0) = \Phi\big(\frac{a-\mu}{\sqrt{\sigma_0^2/n}}\big)$.
Type II error: P(Conclude $H_0|H_1) = 1 - \Phi\big(\frac{a-\mu}{\sqrt{\sigma_0^2/n}}\big)$.
Power Function: P(Conclude H_1) $= \Phi\big(\frac{a-\mu}{\sqrt{\sigma_0^2/n}}\big)$.

Inferences for the Variance

The formula for the confidence interval for variance is as follows:

$$\left[\frac{(n-1)s^2}{\chi^2_{(1+\gamma)/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{(1-\gamma)/2,n-1}}\right]$$

where s is our sample variance, n is our sample size, and γ is our confidence level.

Length of interval

If we want to decrease the size of our interval for our confidence interval, we can increase the population size. In a hypothetical situation, let's say you want a target size of 2δ. Then, in order to determine the number of samples you would need, you would use the equation

$$n \geq \sigma_0^2 \left(\frac{z_{(1+\gamma)/2}}{\delta}\right)^2.$$

If your sample is binomial, you would then use the equation

$$n \geq \frac{1}{4} \left(\frac{z_{(1+\gamma)/2}}{\delta}\right)^2.$$

Finding the more powerful way

Recall that the power function is one minus the Type II error. If we test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at α significance level, then we have that the power function is

$$\beta(\mu) = \Phi\bigg(- \frac{\mu_0 - \mu}{\sqrt{\sigma_0^2/n}} - z_{(1-\alpha)/2}\bigg) + \Phi\bigg(\frac{\mu_0 - \mu}{\sqrt{\sigma_0^2/n}} - z_{(1-\alpha)/2}\bigg).$$

Note that if we have a preselected β_0 then we can derive the formula

$$n \geq \sigma_0^2 \bigg(\frac{z_{1-\beta_0} + z_{(1-\alpha)/2}}{\mu_0 - \mu}\bigg)^2$$

which guarantees that $\beta(\mu) \leq \beta_0$ at μ.
Consider the case where $x_1, \ldots, x_n \sim N(\mu, \sigma_0^2)$ with a known σ_0^2. Consider the test for

$$H_0 : \mu \leq \mu_0 \leftrightarrow H_1 : \mu > \mu_0.$$

We can then derive the formula

$$n \geq \sigma_0^2 \bigg(\frac{z_{1-\beta_0} + z_{1-\alpha}}{\mu - \mu_0}\bigg)^2.$$

Using this, we are able to determine how many samples we need in order to have at most a certain power.

Moment Estimator

Definition 3.0.10. (Theoretical moment)We say that $E(X^k)$ is the k-th theoretical moment about the origin, and we say $E\big((X - \mu)^k\big)$ is the k-th theoretical moment about the mean.

Definition 3.0.11. (Sample moment) We say that $M_k = \frac{1}{n}\sum_{i=1}^n X_i$ is the k-th sample moment about the origin, and we say $M_k^* = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{x})^k$ is the k-th sample moment about the mean.

The general idea for the method is that we equate the first sample moment $m_1 = \bar{x}$ to the theoretical moment $E(X)$, and we set the second sample moment $m_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$ to the second theoretical moment $E(X^2)$, and we continue doing this until we have a series of equations for which we can solve.

Fisher Information

Definition 3.0.12. (Fischer information) The observed Fisher information is given by

$$\hat{I}(s) = - \frac{\partial^2 l(\theta|s)}{\partial \theta^2}\bigg|_{\theta = \hat{\theta}(s)}$$

where $\hat{\theta}(s)$ is the MLE.

Theorem 11. *If certain nice qualities are satisfied (which are just regularity conditions that presumably will always be satisfied) then you have*

$$I(\theta) = Var(S(\theta|X)) = E_\theta\bigg(- \frac{\partial^2 l(\theta|X)}{\partial \theta^2}\bigg).$$

Remark. *Note the X in the function $l(\theta|X)$. This means you only have to find the loglikelihood function for a single value x rather than a sampling of i.i.d $x_1, \ldots, x_n$. This can simplify a lot of calculation.*

Corollary 11.1. *Under i.i.d sampling from a model with Fisher information $I(\theta)$, the Fisher information for a sample of size n is given by $nI(\theta)$.*

Remark. *Using the Theorem and the Corollary, we can find the Fisher information for one point and simply multiply the result by n, rather than having to do the calculation for an i.i.d sampling of $x_1, \ldots, x_n$.*

Theorem 12. *Let* $x_1, \ldots, x_n$ *be i.i.d* $f_\theta(x)$. *Let*

$$I(\theta) = E_\theta\left[\left(\frac{\partial log(f_\theta(x))}{\partial\theta}\right)^2\right] = -E_\theta\left[\frac{\partial^2 log(f_\theta(x))}{\partial\theta^2}\right]$$

be the Fisher information (the variance of the MLE). Then the MLE $\hat{\theta}$ *satisfies*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, I^{-1}(\theta_0)).$$

Theorem 13. *(Delta Theorem) Let* g *be a smooth function (it has continuous derivatives). If* θ *is univariate, then*

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \xrightarrow{D} N(0, I^{-1}(\theta)[g'(\theta)]^2).$$

If θ *is multivariate, we have*

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \xrightarrow{D} N(0, \triangledown^T g(\theta) I^{-1}(\theta) \triangledown g(\theta)).$$

Chapter 4

Bayesian Inference

Remark. *It is important to remember that the probabilities prescribed by the prior represent beliefs. They do not in general correspond to long-run frequencies, although they could in certain circumstances. These models generally come from Statisticians investigations, and in Chapter 9 one would generally explore model checking. We assume for this chapter that all the ingredients make sense, but in application one should take care to ensure that these must be checked if the inferences taken are to be meaningful.*

Definition 4.0.1. (Prior Predictive Distribution) If the data is continuous, we define the prior predictive distribution to be

$$m(s) = \int_{\Omega} \pi(\theta) f_\theta(s) d\theta.$$

If the data is discrete, we replace the integral by a sum to get

$$m(s) = \sum_{\Omega} \pi(\theta) f_\theta(s).$$

Definition 4.0.2. (Prior Density) Since θ is a random variable, we provide a prior density for $\theta \in \Theta$, say $\pi(\theta)$. If

$$\int_{\Theta} \pi(\theta) d\theta = 1,$$

then the prior is proper. If

$$\int_{\Theta} \pi(\theta) d\theta = \infty,$$

then the prior is improper.

Definition 4.0.3. (Posterior Distribution) The posterior distribution of θ is the conditional distribution of θ, given s. The posterior density, or posterior probability function (whichever is relevant) is given by

$$\pi(\theta|s) = \frac{\pi(\theta) f_\theta(s)}{m(s)}.$$

Definition 4.0.4. (MSE) Let $\hat{\theta} = \delta(x)$ be an estimator of θ. The Bayesian method evaluates the mean square error (MSE) based on $L(\delta, \theta) = (\delta - \theta)^2$. The exact definition of Bayesian MSE is

$$E[L(\delta, \theta)] = E[(\delta - \theta)^2].$$

Theorem 14. *The best Bayesian estimator $\delta(x)$, which minimizes the Bayesian MSE, is $\delta(x) = E(\theta|x)$, which is the posterior mean.*

Credible Intervals

Assuming there is one maximum of $q(\theta|x)$. The $(1-\alpha)$ Bayesian credible interval $[l, u]$ for θ satisfies

$$\int_l^u q(\theta|x) = 1 - \alpha$$

and

$$\{\theta \in [l, u]\} = \{q(\theta|x) \geq c\}$$

for some $c > 0$.

Hypothesis Testing and Bayes Factors

To test

$$\psi(\theta) = \psi_0,$$

one can study the posterior probability

$$Q(\psi(\theta) = \psi_0|x).$$

For example, one can test

$$H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$$

by looking at

$$\int_{-\infty}^{\theta_0} q(\theta|x) d\theta.$$

Definition 4.0.5. (Bayes Factors) The Bayesian factor BF_{H_0} in favor of the hypothesis $H_0 : \phi(\theta) = \phi_0$ is defined to be the ratio of the posterior odds in favor of H_0,

$$BF_{H_0} = \left(\frac{Q(\psi(\theta) = \psi_0|x)}{1 - Q(\psi(\theta) = \psi_0|x)}\right) \Big/ \left(\frac{\Pi(\psi(\theta) = \psi_0)}{1 - \Pi(\psi(\theta) = \psi_0)}\right).$$

Remark. *I believe he said Credible Intervals and Hypothesis Testing and Bayes Factors would not be on the final.*

Chapter 5

Homework Exercises and Solutions

Review

Homework 1

Question. (2.6.1) Let $X \sim \text{Uniform}[L, R]$. Let $Y = cX+d$, where $c > 0$. Prove that $Y \sim \text{Uniform}[cL+d, cR+d]$.

Solution. Note that the PDF of X is

$$f(x) = \frac{1}{R-L}, \text{ where } L < x < R$$

and the CDF of X is

$$F(X) = \frac{x-L}{R-L}, \text{ where } L < x < R.$$

Therefore, the CDF of Y is

$$F_Y(y) = P(Y \le y) = P(cX + d \le y) = P\Big(X \le \frac{y-d}{c}\Big) = \frac{y-(cL+d)}{(cR+d)-(cL+d)}.$$

Thus, we have that $Y \sim \text{Uniform}[cL+d, cR+d]$.

Question. (2.6.3) Let $X \sim N(\mu, \sigma^2)$. Let $Y = cX + d$, where $c > 0$. Prove that $Y \sim N(c\mu + d, c^2\sigma^2)$.

Solution. The CDF of X is

$$F(X) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(t-\mu)^2}{2\sigma^2}}\, dt.$$

The CDF of Y is

$$\begin{aligned} F_Y(y) &= P(cX + d \le y) = P(X \le \frac{y-d}{c}) \\ &= \int_{-\infty}^{(y-d)/c} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(t-\mu)^2}{2\sigma^2}}\, dt \\ &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}c} e^{\frac{-[t-(c\mu+d)]^2}{2c^2\sigma^2}}\, dt, \end{aligned}$$

where $s = (t-d)/c$. Therefore, $Y \sim N(c\mu + d, c^2\sigma^2)$.

Question. (2.6.4) Let $X \sim \text{Exponential}(\lambda)$. Let $Y = cX$, where $c > 0$. Prove that $Y \sim \text{Exponential}(\lambda/c)$.

Solution. Note that the CDF of X is

$$F(x) = e^{-\lambda x}, \text{ where } x > 0.$$

The CDF of $Y = cX$ for a positive c is

$$F_Y(y) = P(cX \le y) = P(X \le y/c) = e^{-\lambda y/c} = e^{-(\lambda/c)y}, \text{ where } y > 0.$$

Therefore, $Y \sim \text{Exponential}(\lambda/c)$.

Question. (2.7.9) Let X and Y have joint density $f_{X,Y}(x,y) = (x^2+y)/4$ for $0 < x < y < 2$, otherwise $f_{X,Y}(x,y) = 0$. Compute each of the following.
(a) The marginal density $f_X(x)$ for all $x \in \mathbb{R}$.
(b) The marginal density $f_Y(y)$ for all $y \in \mathbb{R}$.
(c) $P(Y < 1)$.

Solution. (a) We must integrate over all values of y. Hence, we have

$$\frac{1}{4}\int_x^2 (x^2+y)dy = \frac{1}{4}\left(x^2y + \frac{y^2}{2}\right)\Big|_x^2 = -\frac{x^3}{4} + \frac{3x^2}{8} + \frac{1}{2} = f_X(x).$$

(b) Analogously, we integrate over all values of x. Hence, we have

$$\frac{1}{4}\int_0^y (x^2+y)dx = \frac{1}{4}\left(\frac{x^3}{3} + xy\right)\Big|_0^y = \frac{y^3}{12} + \frac{y^2}{4}.$$

(c) Since the function is continuous, we have $P(Y < 1) = P(Y \leq 1)$. Using part (b), we integrate from 0 to 1. Hence,

$$\int_0^1 \left(\frac{y^3}{12} + \frac{y^2}{4}\right)dy = \frac{5}{48}.$$

Question. (2.8.8) Let X and Y be jointly absolutely continuous random variables. Suppose $X \sim$ Exponential(2) and that $P(Y > 5|X = x) = e^{-3x}$. Compute $P(Y > 5)$.

Solution. We have $P(Y > 5|X = x) = e^{-3x}$, then note that $P_X(x) = 2e^{-2x}$, and so $P(Y > 5|X = x) = \frac{P_{X,Y}(x,y>5)}{2e^{-2x}} = e^{-3x} \to 2e^{-5x} = P(X, Y > 5)$ and so we have $P(Y > 5) = \int_0^\infty P(X, Y > 5) = 2/5$.

Remark. *Note that his notes ask for $P(Y = 5)$, which seems to be a typo. He does not have the same answer as me. He wrote*

$$P(Y = 5) = \int_0^\infty P(Y = 5|X = x)2e^{-2x}dx = 0.4.$$

Question. (2.8.12) Suppose that $X \sim$ Bernoulli(1/3) and $Y \sim$ Poisson(λ), with X and Y independent and with $\lambda > 0$. Compute $P(X = 1|Y = 5)$.

Solution. Note that we have X and Y independent, and so $P(X = 1|Y = 5) = P(X = 1) = 1/3$.

Homework 2

Question. (3.1.4) Let $X \sim \text{Bernoulli}(\theta_1)$ and $Y \sim \text{Binomial}(n, \theta_2)$. Compute $E(4X - 3Y)$.

Solution. By linearity, we have $E(4X - 3Y) = 4E(X) - 3E(Y)$. Using the properties of the distributions, we have $E(X) = \theta_1$ and $E(Y) = \theta_2$, and so $E(4X - 3Y) = 4\theta_1 - 3\theta_2$.

Remark. *This is the answer off of his notes. Since $Y \sim$ Binomial(n, θ_2), then by the table one sees that $E(Y) = n\theta_2$. Hence, we have $E(4X - 3Y) = 4\theta_1 - 3n\theta_2$.*

Question. (3.1.6) Let $Y \sim \text{Binomial}(100, 0.3)$ and $Z \sim \text{Poisson}(7)$. Compute $E(Y + Z)$.

Solution. Note $E(X) = 100 * 0.3 = 30$ and $E(Z) = 7$. Using the linearity again, we have $E(X + Z) = E(X) + E(Z) = 30 + 7 = 37$.

Question. (3.2.5) Let $X \sim \text{Uniform}[3, 7]$ and $Y \sim \text{Exponential}(9)$. Compute $E(-5X - 6Y)$.

Solution. Note $E(X) = 5$ and $E(Y) = 1/9$, and so by the linearity of expectation we have $E(-5X - 6Y) = -5E(X) - 6E(Y) = -77/3$.

Question. (3.2.7) Let $Y \sim \text{Exponential}(9)$ and $Z \sim \text{Exponential}(8)$. Compute $E(Y + Z)$.

Solution. Note that $E(Y) = 1/9$ and $E(Z) = 1/8$, and so by linearity of expectation we have $E(Y + Z) = E(Y) + E(Z) = 1/9 + 1/8 = 17/72$.

Question. (3.3.3) Let X and Y have joint density

$$f_{X,Y}(x, y) = 4x^2y + 2y^5 \text{ when } 0 \le x \le 1, 0 \le y \le 1.$$

Compute Corr(X,Y).

Solution. Alright, here comes a bunch of integrals. I'm not going to show much work here.

$$\begin{aligned}
E(X) &= \int_0^1 \int_0^1 x(4x^2y + 2y^5)dxdy = 2/3 \\
E(X^2) &= \int_0^1 \int_0^1 x^2(4x^2y + 2y^5)dxdy = 23/45 \\
E(Y) &= \int_0^1 \int_0^1 y(4x^2y + 2y^5)dxdy = 46/63 \\
E(Y^2) &= \int_0^1 \int_0^1 y^2(4x^2y + 2y^5)dxdy = 7/12 \\
E(XY) &= \int_0^1 \int_0^1 xy(4x^2y + 2y^5)dxdy = 10/21.
\end{aligned}$$

Therefore, we can note that $V(X) = 23/45 - (2/3)^2 = 0.0667$, $V(Y) = 7/12 - (46/63)^2 = 0.0575$, and $\text{Cov}(X, Y) = 10/21 - (2/3)(46/63) = -0.01058$, and so

$$\text{Corr}(X, Y) = \frac{-0.01058}{\sqrt{(0.0667)(0.0575)}} = -0.1708.$$

Question. (3.3.6) Let X, Y, and Z be three random variables, and suppose that X and Z are independent. Prove that $\text{Cov}(X + Y, Z) = \text{Cov}(Y, Z)$.

Solution. By a theorem, we have $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$. Note that $\text{Cov}(X, Z) = 0$ since they are independent. Hence, $\text{Cov}(X + Y, Z) = \text{Cov}(Y, Z)$.

Question. (3.3.7) Let $X \sim \text{Exponential}(3)$ and $Y \sim \text{Poission}(5)$. Assume that X and Y are independent, and let $Z = X + Y$.
(a) Compute $\text{Cov}(X, Z)$.
(b) Compute $\text{Corr}(X, Z)$.

Solution. (a) Note that $E(X) = 1/3$, $V(X) = 1/9$, $E(Y) = V(Y) = 5$. Then since $V(X) = E(X^2) - E(X)^2$, we have $E(X^2) = V(X) + E(X)^2 = 2/9$, and likewise $E(Y^2) = V(Y) + E(Y)^2 = 30$. Therefore, $E(Z) = 16/3$ by linearity of expectance. Note that $E(XZ) = E(X(X+Y)) = E(X^2) + E(X)E(Y) = 17/9$. Therefore, $\mathrm{Cov}(X, Z) = 17/9 - (16/3)(1/3) = 1/9$.
(b) $E(Z^2) = E(X+Y)^2 E(X^2 + 2XY + Y^2) = 2/9 + 2(1/3)(5) + 30 = 302/9$. Substituting this into the formula grants us

$$\mathrm{Corr}(X, Z) = \frac{1/9}{\sqrt{(1/9)(302/9)}} = 0.0575.$$

Chapter 6

Homework 3

Question. (6.1.3) Suppose that the lifelengths (in thousands of hours) of light bulbs are distributed Exponential(θ), where $\theta > 0$ is unknown. If we observe $\bar{x} = 5.2$ for a sample of 20 light bulbs, record a representative likelihood function. Why is it that we only need to observe the sample average to obtain a representative likelihood?

Solution. Note that, since we have Exponential(θ), then the function is $f_\theta(x) = \theta e^{-\theta x}$. Therefore, the likelihood function is

$$\prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i} = \theta^n e^{-(n\theta)\bar{x}},$$

and therefore $\bar{x}$ is a sufficient statistic by the factorization theorem.

Question. (6.1.6) Suppose that $(x_1, \ldots, x_n)$ is a sample from a Bernoulli(θ) distribution, where $\theta \in [0, 1]$ is unknown. Determine the likelihood function and a minimal sufficient statistic for this model.

Solution. Using the same reasoning as prior, one can find that the likelihood function is

$$L(\theta|s) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}.$$

Since the dimension of $\sum_{i=1}^{n} x_i$ is one, we have that it is a minimal sufficient statistic by a theorem from the class notes. One can also use the fact that $\bar{x} = \sum_{i=1}^{n} x_i/n$ to show that $\bar{x}$ is a minimal sufficient statistic as well.

Remark. *I flip between using the notation of $L(\theta|s)$ and $L(\theta)$. I'm pretty sure it's safe to assume that these functions are equivalent throughout the rest of the notes.*

Question. (6.1.7) Suppose $(x_1, \ldots, x_n)$ is a sample from a Poisson(θ) distribution, where $\theta > 0$ is unknown. Determine the likelihood function and a minimal sufficient statistic for this model.

Solution. Using the same reasoning as prior, one finds that

$$L(\theta|s) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \frac{e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \theta^{\sum_{i=1}^{n} x_i}.$$

Also by the same reasoning as prior, we have $\sum_{i=1}^{n} x_i$ is a sufficient statistic, as well as $\bar{x}$.

Question. (6.1.11) Suppose that we have a statistical model $\{f_\theta : \theta \in [0, 1]\}$, and we observe x_0. Is it true that $\int_0^1 L(\theta|x_0) d\theta = 1$? Explain why or why not.

Solution. Simply remark that $f_\theta(x)$ is a PDF of x but not of θ.

Question. (6.1.12) Suppose that $(x_1, \ldots, x_n)$ is a sample from a Geometric(θ) distribution, where $\theta \in [0, 1]$ is unkown. Determine the likelihood function and a minimal sufficient statistic for this model.

Solution. Recall that the PMF for a Geometric distribution is $p_X(x) = \theta^x(1-\theta)$. Therefore, the likelihood function is

$$L(\theta|s) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta) = (1-\theta)^n \theta^{\sum_{i=1}^{n} x_i}.$$

Hence, $\bar{x}$ and $\sum_{i=1}^{n} x_i$ are sufficient statistics.

Question. (6.1.14) Suppose that one statistician records a likelihood function as θ^2 for $\theta \in [0, 1]$ while another statistician records a likelihood function as $100\theta^2$ for $\theta \in [0, 1]$. Explain why these likelihood functions are effectively the same.

Solution. Note that they only differ by a constant, and so by a theorem we have that they are essentially equivalent.

Question. (6.1.22) For the location-scale normal model, establish that the point where the likelihood is maximized is given by $(\bar{x}, \hat{\sigma}^2)$ as defined in Example 6.1.8.

Solution. It is sufficient to show that the first-order derivative is zero and the second-order derivative is negative. We have then that

$$L(\mu, \sigma^2|s) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

(There are a few steps in between of algebra.)

$$= \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{n}{2\sigma^2}(\hat{\sigma}^2+(\bar{x}-\mu)^2)^2}.$$

Note then that the loglikelihood function is

$$l(\mu, \sigma^2|s) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{n}{2\sigma^2}(\hat{\sigma}^2 + (\bar{x}-\mu)^2).$$

The first-order partial derivatives are

$$\frac{\partial l(\mu, \sigma^2|s)}{\partial \mu} = \frac{n}{\sigma^2}(\bar{x}-\mu)$$

and

$$\frac{\partial l(\mu, \sigma^2|s)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{n\hat{\sigma}^2}{2(\sigma^2)^2}.$$

Then we obtain

$$l(\mu, \sigma^2|s) = logL(\mu, \sigma^2|s)\Big|_{\mu=\bar{x}} = 0$$

and

$$\frac{\partial l(\mu, \sigma^2|s)}{\partial(\sigma^2)}\Big|_{\sigma^2=\hat{\sigma}^2} = 0.$$

Taking the second-order partial derivative follows similarly and we see that these values are maximum (note we could also use Dr. Zhang's theorem from class to force this fact). Therefore, $(\bar{x}, \hat{\sigma}^2)$ is a maximizer.

Homework 4

Question. (6.2.3) If $(x_1, \ldots, x_n)$ is a sample from a Bernoulli(θ) distribution, where $\theta \in [0,1]$ is unknown, then determine the MLE of θ^2.

Solution. One could derive this fact, or just note that since MLE of θ is $\bar{x}$, then the MLE of θ^2 is $\bar{x}^2$.

Question. (6.2.4) If $(x_1, \ldots, x_n)$ is a sample from Poission(θ) distribution, where $\theta \in (0, \infty)$ is unknown, then determine the MLE of θ.

Solution. Note that the loglikelihood function

$$l(\theta) = \log\left(\prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta}\right)$$

$$= \log\left(\frac{1}{\prod_{i=1}^{n} x_i! \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}\right)$$

$$= -\sum_{i=1}^{n} \log(x_i!) - n\bar{x}\log(\theta) + n\theta.$$

Then,

$$l'(\theta) = -\frac{n\bar{x}}{\theta} + n \to \hat{\theta} = \bar{x}.$$

Question. (6.2.5) If $(x_1, \ldots, x_n)$ is a sample from a Gamma(α_0, θ) distribution, where $\alpha_0 > 0$ and $\theta \in (0, \infty)$ is unknown, then determine the MLE of θ.

Solution. Note that the PDF is

$$f_\theta(x) = \frac{\theta^{\alpha_0} x^{\alpha_0 - 1}}{\Gamma(\alpha_0)} e^{-\theta x},$$

where α_0 is known. The loglikelihood function is then

$$l(\theta) = \log\left(\prod_{i=1}^{n} \frac{\theta^{\alpha_0} x_i^{\alpha_0 - 1}}{\Gamma(\alpha_0)} e^{-\theta x_i}\right)$$

$$= n\alpha_0 \log(\theta) - n\log(\Gamma(\alpha_0)) + (\alpha_0 - 1)\sum_{i=1}^{n} \log(x_i) - n\theta\bar{x}.$$

Note then that

$$l'(\theta) = \frac{n\alpha_0}{\theta} - n\bar{x} \to \hat{\theta} = \frac{\alpha_0}{\bar{x}}.$$

Question. (6.2.6) Suppose that $(x_1, \ldots, x_n)$ is the result of independent tosses of a coinwhere we toss until the first head occurs and where the probability of a head on a single toss is $\theta \in [0, 1)$. Determine the MLE of θ.

Solution. Notice that we have a geometric distribution, and so the PMF is $f_\theta(x) = (1-\theta)^{x-1}\theta$. The loglikelihood function is then

$$l(\theta) = \log\left(\prod_{i=1}^{n} (1-\theta)^{x_i - 1}\theta\right) = n(\bar{x} - 1)\log(1-\theta) + n\log(\theta).$$

We then take the derivative and get

$$l'(\theta) = -\frac{n(\bar{x} - 1)}{1-\theta} + \frac{n}{\theta} \to \hat{\theta} = \frac{1}{\bar{x}}.$$

Question. (6.2.9) If $(x_1, \ldots, x_n)$ is a sample from a Pareto(α) distribution, where $\alpha > 0$ is unkown, then determine the MLE of α.

Solution. Recall that the PDF is $f_\alpha(x) = \alpha(1+x)^{-\alpha-1}$. The loglikelihood function is then

$$l(\alpha) = \log\biggl(\prod_{i=1}^n [\alpha(1+x_i)]^{-\alpha-1}\biggr) = n\log(\alpha) - (\alpha+1)\sum_{i=1}^n \log(1+x_i).$$

Then taking the derivative and setting it equal to 0, we have

$$l'(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^n \log(1+x_i) \to \hat{\alpha} = \frac{n}{\sum_{i=1}^n \log(1+x_i)}.$$

Question. (6.2.10) If $(x_1, \ldots, x_n)$ is a sample from a log-normal(τ) distribution, where $\tau > 0$ is unknown, then determine the MLE of τ.

Solution. Note that the PDF is

$$f_\tau(x) = \frac{1}{\sqrt{2\pi\tau^2}} exp\biggl(-\frac{\log^2(x)}{2\tau^2}\biggr)\frac{1}{x}.$$

The loglikelihood function is then

$$l(\tau) = \log\biggl(\prod_{i=1}^n \biggl[\frac{1}{\sqrt{2\pi}\tau} exp\biggl(-\frac{\log^2(x_i)}{2\tau^2}\biggr)\frac{1}{x_i}\biggr]\biggr)$$

$$= -\frac{n}{2}\log(2\pi) - \sum_{i=1}^n \log(x_i) - n\log(\tau) - \frac{1}{2\tau^2}\sum_{i=1}^n \log^2(x_i).$$

Taking the derivative and setting it equal to 0 grants us

$$l'(\tau) = -\frac{n}{\tau} + \frac{1}{\tau^3}\sum_{i=1}^n \log^2(x_i) \to \hat{\tau} = \biggl(\frac{1}{n}\sum_{i=1}^n \log^2(x_i)\biggr)^{1/2}.$$

Homework 5

Question. (6.2.12) If $(x_1, \dots, x_n)$ is a sample from an $N(\mu_0, \sigma^2)$ distribution, where $\sigma^2 > 0$ is unkown and μ_0 is known, then determine the MLE of σ^2 How does this MLE differ from the plug-in MLE of σ^2 computed using the location-scale normal model?

Solution. The loglikelihood function is

$$l(\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

Then

$$l'(\sigma^2) = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu_0)^2 \to \hat{\sigma}^2(\mu_0) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

Comparing it with

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

we see that they are clearly not identical.

Question. (6.2.13) Explain why it is not possible that the function $\theta^3 exp(-(\theta - 5.3)^2)$ for $\theta \in \mathbb{R}$ is a likelihood function.

Solution. Note that it's values can be negative for $\theta < 0$, and so we have that it cannot be a likelihood function.

Question. (6.2.21) If $(x_1, \dots, x_n)$ is a sample from an $N(\mu, 1)$ distribution where $\mu \geq 0$ is unknown, determine the MLE of μ.

Solution. The loglikelihood function is

$$l(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}[n(\bar{x} - \mu)^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2].$$

Hence, we see it is maximized at $\mu = \bar{x}$ if $\bar{x} > 0$ and 0 if $\bar{x} = 0$. So we have that the MLE is $\max(0, \bar{x})$.

Question. (6.2.22) Prove that, if $\hat{\theta}(s)$ is the MLE for a mdoel for response s and if T is a sufficient statistic for the model, then $\hat{\theta}(s)$ is also the MLE for the model $T(s)$.

Solution. For the model $\mathbf{x} = (x_1, \dots, x_n)$, the likelihood function is

$$L(\theta|x) = h(\mathbf{x})g_\theta(T).$$

If the model for T is used such that $T(x_1) = T(x_2)$, then $L(\theta|x_1) = h(x_1)g_\theta(T)$ and $L(\theta|x_2) = h(x_2)g_\theta(T)$, indicating they have the same maximum. Therefore, $\hat{\theta}$ is also the MLE for the model T.

Question. (6.2.24) If $(x_1, \dots, x_n)$ is a sample from a Uniform$[\theta_1, \theta_2]$ distribution with

$$\Omega = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 < \theta_2\},$$

determine the MLE of (θ_1, θ_2).

Solution. Note that the PDF is

$$f_\theta(x) = \frac{1}{\theta_2 - \theta_1}I(\theta_1 \leq x \leq \theta_2).$$

The likelihood function is

$$l(\theta) = \frac{1}{\theta_2 - \theta_1}I(\theta_1 \leq \min_{i \leq n}(x_i) \leq \max_{i \leq n}(x_i) \leq \theta_2).$$

To make $l(\theta)$ large, we need to increase θ_1 and decrease θ_2. Hence, we have $\theta_1 = \min_{i \leq n}$ and $\theta_2 = \max_{i \leq n}$.

Homework 6

Question. (6.3.2) Suppose measurements (in centimeters) are taken using an instrument. There is error in the measuring process, and a measurement is assumed to be distributed $N(\mu, \sigma_0^2)$, where μ is the exact measurement and $\sigma_0^2 > 0$ is unknown. If the (n=10) measurements 4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3 were obtained, assess the hypothesis $H_0 : \mu = 5$ and compute a 0.95-confidence interval for μ.

Solution. We obtain $\bar{x} = 4.88$ and $s^2 = 0.484$. Thus, the test statistic is

$$t = \left| \frac{\bar{x} - 5}{\sqrt{s^2/n}} \right| = \left| \frac{4.88 - 5}{\sqrt{0.484/10}} \right| = 0.5455$$

which is less than $t_{0.975,9} = 2.262$. Therefore, we fail to reject H_0 and conclude $\mu = 5$. The 0.95-confidence interval for μ is

$$\bar{x} \pm t_{0.975,9} \frac{s}{\sqrt{10}} = 4.88 \pm 2.262 \sqrt{\frac{0.484}{10}} = [4.38, 5.38].$$

Question. (6.3.3) Marks on an exam in a statistics course are assumed to be normally distributed with unknown mean but with variance equal to 5. A sample of four students is selected, and their marks are 52, 63, 64, 84. Assess the hypothesis $H_0 : \mu = 60$ by computing the relevant P-value and compute a 0.95-confidence interval for the unknown μ.

Solution. We obtain $\bar{x} = 65.75$. Note that we have $\sigma_0^2 = 5$, and so

$$z = \left| \frac{\bar{x} - 60}{\sqrt{\sigma_0^2/n}} \right| = \left| \frac{65.75 - 60}{\sqrt{5/4}} \right| = 5.14.$$

The P-value is $2\Phi(-5.14) = 0$. Thus, we reject H_0. The 0.95-confidence interval for μ is

$$\bar{x} \pm 1.96 \sqrt{\frac{5}{4}} = [63.56, 67.94].$$

Question. (6.3.4) Suppose that in Exercise 6.3.3 we drop the assumption that the population variance is 5. Assess the hypothesis $H_0 : \mu = 60$ by computing the relevant P-value and compute a 0.95-confidence interval for the unknown μ.

Solution. Note that $s^2 = 177.58$, and so we have

$$t = \left| \frac{\bar{x} - 60}{\sqrt{s^2/n}} \right| = \left| \frac{65.75 - 60}{\sqrt{177.58/4}} \right| = 0.8630.$$

We have then that the P-value is $2T_3(-0.863) = 0.4516$. Thus, we accept H_0 and conclude $\mu = 60$. The 0.95-confidence interval is

$$\bar{x} \pm 3.182 \sqrt{\frac{177.58}{4}} = [44.55, 86.95].$$

Question. (6.3.6) Assume that the speed of light data in the following table is a sample from an $N(\mu, \sigma^2)$ distribution for some unknown values of μ and σ^2. Determine a 0.99-confidence interval for μ. Assess the null hypothesis $H_0 : \mu = 24$.

28	26	33	24	34	-44	27	16	40	-2	29
22	24	21	25	30	23	29	31	19	24	20
36	32	36	28	25	21	28	29	37	25	28
26	30	32	36	26	30	22	36	23	27	27
28	27	31	27	26	33	26	32	32	24	39
28	24	25	32	25	29	27	28	29	16	23

Solution. We obtain $\bar{x} = 26.21$ and $s^2 = 115.46$. The 0.99-confidence interval is

$$\bar{x} \pm t_{0.995,65}\sqrt{\frac{s^2}{66}} = 26.21 \pm 2.654\sqrt{\frac{115.46}{66}} = [22.7, 29.72].$$

Since the 0.99-confidence interval contains 24 at the 0.01 significance level, we fail to reject the null hypothesis.

Question. (6.3.8) A polling firm conducts a poll to determine what proportion θ of voters in a given population will vote in an upcoming election. A random sample of $n = 250$ was taken from the population, and the proportion answering yes was 0.62. Assess the hypothesis $H_0 : 0.65$ and construct an approximate 0.90-confidence interval for θ.

Solution. The MLE is $\hat{\theta} = 0.62$. The test statistic is

$$z = \left|\frac{0.62 - 0.65}{\sqrt{0.65(1 - 0.65)/250}}\right| = 0.9945.$$

Then, we accept $H_0 : p = 0.65$. In addition, we may use

$$z = \left|\frac{0.62 - 0.65}{\sqrt{0.62(1 - 0.62)/250}}\right| = 0.9772.$$

The conclusion is consistent with the previous one. You only need to do one of these. The 0.90-confidence interval for θ is

$$0.62 \pm 1.645\sqrt{\frac{0.62(1 - 0.62)}{250}} = [0.5695, 0.6705].$$

Question. (6.3.9) A coin was tossed $n = 1000$ times, and the proportion of heads observed was 0.51. Do we have evidence to conclude the coin is unfair?

Solution. Assume $X \sim Bin(1000, \theta)$. The MLE is $\hat{\theta} = X/n = 0.51$. We test $H_0 : \theta = 0.5$ against $H_1 : \theta \neq 0.5$. The test statistic is

$$t = \left|\frac{0.51 - 0.5}{0.5(1 - 0.5)/1000}\right| = 0.6324 < 1.96.$$

Thus, we accept $H_0 : \theta = 0.5$ at the 0.05 significance level. One could also instead use

$$t = \left|\frac{0.51 - 0.5}{0.51(1 - 0.51)/1000}\right| = 0.6326 < 1.96.$$

Question. (6.3.10) How many times must we toss a coin to ensure that 0.95-confidence interval for the probability of heads on a single toss has length less than 0.1, 0.05, and 0.01, respectively?

Solution. We use the formula

$$n \geq \left(\frac{1.96}{2\delta}\right)^2$$

where the length of the interval is 2δ. We have then $\delta = 0.05, 0.025, 0.005$ and obtain $n = 385, n = 1537$, and $n = 38416$, respectively.

Homework 7

Question. (6.3.12) Suppose a measurement on a population is assumed to be distributed $N(\mu, 2)$ where $\mu \in \mathbb{R}$ is unknown and that the size of the population is very large. A researcher wants to determine a 0.95-confidence interval for μ that is no longer than 1. What is the minimum sample size that will guarantee this?

Solution. The length of the 0.95-confidence interval is $2(1.96)\sigma_0/\sqrt{n} = 5.544/\sqrt{n}$. If we need $5.544/\sqrt{n} \le 1$, then $n \ge (5.544/1)^2 = 30.7$. Thus, we choose $n = 31$.

Question. (6.3.13) Suppose $(x_1, \ldots, x_n)$ is a sample from Bernoulli(θ), with $\theta \in [0, 1]$ unknown.
(a) Show that $\sum_{i=1}^{n}(x_i - \bar{x})^2 = n\bar{x}(1 - \bar{x})$.
(b)If $X \sim$ Bernoulli(θ), then $\sigma^2 = Var(X) = \theta(1 - \theta)$. Record the relationship between the plug-in estimate of σ^2 and that given by s^2.
(c) Since s^2 is an unbiased estimator of σ^2, use the results in part (b) to determine the bias in the plug-in estimate. What happens to this bias as $n \to \infty$?

Solution. (a) This is just a series of manipulations. We have

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} x_i - n\bar{x}^2 = n\bar{x} - n\bar{x}^2 = n\bar{x}(1 - \bar{x}).$$

(b) The plug-in estimate of σ^2 is $\hat{\sigma}^2 = \bar{x}(1 - \bar{x})$. The estimator of σ^2 based s^2 is

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n\hat{\theta}(1 - \hat{\theta})}{n-1} = \frac{n\hat{\sigma}^2}{n-1}.$$

(c) We have

$$E(\hat{\sigma}^2) = \frac{n-1}{n}E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 = \frac{(n-1)\theta(1-\theta)}{n}.$$

Thus,

$$Bias(\hat{\sigma}^2) = \frac{(n-1)\theta(1-\theta)}{n} - \theta(1-\theta) = \frac{\theta(1-\theta)}{n}.$$

Question. (6.3.14) Suppose you are told that, based on some data, a 0.95-confidence interval for a characteristic $\psi(\theta)$ is given by (1.23, 2.45). You are then asked if there is any evidence against the hypothesis $H_0 : \psi(\theta) = 2$. State your conclusin and justify your reasoning.

Solution. Since the 95% confidence interval contains 2, we conclude $H_0 : \psi(\theta) = 2$ at 0.05 significance level.

Question. (6.3.15) Suppose that x_1 is a value from Bernoulli(θ) with $\theta \in [0, 1]$ unknown.
(a) Is x_1 an unbiased estimator of θ
(b) Is x_1^2 an unbiased estimator of θ^2?

Solution. (a) $E(x_1) = \theta$ and $Var(x_1) = \theta(1 - \theta)$. Thus, x_1 is an unbiased estimator of θ.
(b) $E(x_1^2) = Var(x_1) + E(x_1)^2 = \theta(1 - \theta) + \theta 62 = \theta$. Thus, it is not an unbiased estimator of θ^2.

Question. (6.3.24) Suppose we have two unbiased estimators T_1 and T_2 of $\psi(\theta) \in \mathbb{R}$.
(a) Show that $\alpha T_1 + (1 - \alpha)T_2$ is also an unbiased estimator of $\psi(\theta)$ whenever $\alpha \in [0, 1]$.
(b) If T_1 and T_2 are also independent, then calculate $Var_\theta(\alpha T_1 + (1 - \alpha)T_2)$ in terms of $Var_\theta(T_1)$ and $Var_\theta(T_2)$.
(c) For the situation in part (b), determine the best choice of α in the sense that for this choice $Var_\theta(\alpha T_1 + (1 - \alpha)T_2)$ is smallest. What is the effect on this combined estimator of T_1 having a very large variance relative to T_2?
(d) Repeat parts (b) and (c) but now do not assume T_1 and T_2 are independent, so $Var_\theta(\alpha T_1 + (1 - \alpha)T_2)$ will also involve $Cov_\theta(T_1, T_2)$.

Solution. (a) $E(\alpha T_1 + (1-\alpha)T_2) = \alpha E(T_1) + (1-\alpha)E(T_2) = \psi(\theta)$. Thus, $\alpha T_1 + (1-\alpha)T_2$ is an unbiased estimator for any $\alpha \in [0,1]$.
(b) If T_1 and T_2 are independent, then

$$Var(\alpha T_1 + (1-\alpha)T_2) = \alpha^2 Var(T_1) + (1-\alpha)^2 Var(T_2).$$

(c) Differentiating with respect to α, we obtain

$$\frac{\partial}{\partial\alpha} Var(\alpha T_1 + (1-\alpha)T_2) = 2\alpha Var(T_1) - 2(1-\alpha)Var(t_2).$$

Let

$$2\alpha Var(T_1) - 2(1-\alpha)Var(T_2) = 0.$$

We obtain

$$\alpha_{Best} = \frac{Var(T_2)}{Var(T_1) + Var(T_2)}.$$

If $Var(T_1)$ is much larger than $Var(T_2)$, then α is small indiciating that we should put more weight on T_2.
(d) If T_1 and T_2 are not independent, then

$$Var(\alpha T_1 + (1-\alpha)T_2) = \alpha^2 Var(T_1) + (1-\alpha)^2 Var(T_2) + 2\alpha(1-\alpha)Cov(T_1, T_2).$$

Differentiating α, we obtain

$$\frac{\partial}{\partial\alpha} Var(\alpha T_1 + (1-\alpha)T_2) = 2\alpha Var(T_1) - 2(1-\alpha)Var(T_2) + (2-4\alpha)Cov(T_1, T_2),$$

implying that

$$\alpha_{best} = \frac{Var(T_2) - Cov(T_1, T_2)}{Var(T_1) + Var(T_2) - 2Cov(T_1, T_2)}.$$

Question. (6.3.25) Suppose that $(x_1, \ldots, x_n)$ is a sample from an $N(\mu, \sigma^2)$ distribution, where $\mu \in \mathbb{R}$ is unknown and σ^2 is known. Suppose we want to make inferences about the interval $\psi(\mu) = (-\infty, \mu)$. Consider the problem of finding an interval $C(x_1, \ldots, x_n) = (-\infty, u(x_1, \ldots, x_n))$ that ocovers the interval $(-\infty, \mu)$ with probability at least γ. So we want u such that for every μ,

$$P_\mu(\mu \le u(x_1, \ldots, x_n)) \ge \gamma.$$

Note that $(-\infty, \mu) \subset (-\infty, u(x_1, \ldots, x_n))$ if and only if $\mu \le u(x_1, \ldots, x_n)$, so $C(x_1, \ldots, x_n)$ is called a left-sided γ-confidence interval for μ. Obtain an exact left-sided γ-confidence interval for μ using $u(x_1, \ldots, x_n) = \bar{x} + k(\sigma_0/\sqrt{n})$, i.e., find the k that gives this property.

Solution. We just have a series of equations. Hence

$$\begin{aligned} P(\mu \le \bar{x} + k(\sigma/\sqrt{n})) &= P(\bar{x} \ge \mu - k(\sigma/\sqrt{n}) \\ &= 1 - \Phi\left(\frac{\mu - (\mu - k\sigma/\sqrt{n})}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi(k). \end{aligned}$$

Thus, $k = z_{1-\gamma}$ (classical notation) or $k = z_\gamma$ (textbook notation).

Question. (6.3.26) Suppose that $(x_1, \ldots, x_n)$ is a sample from a $N(\mu, \sigma^2)$ distribution, where μ is unknown and σ^2 is known. Suppose we want to assess the hypothesis $H_0 : \mu \le \mu_0$. Under these circumstances, we say that the observed value $\bar{x}$ is suprising if $\bar{x}$ occurs in a region of low probability for every distribution in H_0. Therefore, a sensible P-value for this problem is $\max_{\mu \in H_0} P_\mu(\bar{X} > \bar{x})$. Show that this leads to the P-value $1 - \Phi((\bar{x} - \mu_0)/(\sigma/\sqrt{n}))$.

Solution. Since $\bar{X} \sim N(\mu, \sigma^2/n)$,

$$\max_{\mu \in H_0} P_\mu(\bar{X} \ge \bar{x}) = \max_{\mu \le \mu_0}(1 - \Phi\left(\frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}\right)) = 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}}\right),$$

where the last equation holds since $1 - \Phi((\bar{x} - \mu)/(\sqrt{\sigma^2/n}))$ is increasing in μ.

Homework 8

Question. (6.4.2) Determine the method of moments estimator of the population variance. Is this estimator unbiased for the population variance? Justify your answer.

Solution. Based on the first-order moment estimation equation, we obtain $\hat{\mu}_{ME} = \bar{x}$. Based on the second-order moment estimation equation, we obtain

$$\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \to \hat{\sigma}^2_{ME} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Since

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

is an unbiased estimator of σ^2, $\hat{\sigma}^2_{ME}$ is a biased estimator of σ^2. Its bias

$$E(\hat{\sigma}^2_{ME} - \sigma^2) = \frac{n-1}{n}E(s^2) - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = \frac{\sigma^2}{n}.$$

Question. (6.4.3) The coefficient of variation for a population measurement with nonzero mean is given by σ/μ, where μ is the population mean and σ is the population standard deviaton. What is the method of moments estimate of the coefficient of variation? Prove tha the coefficient of variation is invariant under rescalings of the distribution, i.e., under transformations of the form $T(x) = cx$ for some constant $c > 0$. It is this invariance that leads to the coefficient of variation being an appropriate measure of sampling variability in certain problems, as it is independent of the units we are for the measurement.

Solution. Based on the first-order moment estimation equation, we obtain $\mu_{ME} = \bar{x}$. Based on the second-order moment estimation equation, we obtaim $\hat{\sigma}^2_{ME} = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$. Therefore, the moment estimator of $\psi = \sigma/\mu$ is $\hat{\psi}_{ME} = (\sum_{i=1}^{n}(x_i - \bar{x})^2/n)^{1/2}/\bar{x}$. Let $y_i = cx_i$ be the value derived under rescalings of the data. Then $\bar{y} = c\bar{x}$ and $\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} = c\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, implying that $\hat{\psi}_{ME}$ based on $y_1, \ldots, y_n$ is still the same as the value based on $x_1, \ldots, x_n$.

Question. (6.4.5) Verify that the third moment of an $N(\mu, \sigma^2)$ distribution is given by $\mu_3 = \mu^3 + 3\mu\sigma^2$. Because the normal distribution is specified by its first two moments, any characteristic of the normal distribution can be estimated by simply plugging in the MLE estimates of μ and σ^2. Compare the method of moments estimator of μ_3 with this plug-in MLE estimator, i.e., determine whether they are the same or not.

Solution. Using the fact that $E[(X - \mu)^3) = 0$, we obtain

$$\begin{aligned} 0 = E[(X - \mu)^3] &= E(X^3 - 3X^2\mu + 3X\mu^2 - \mu^3) \\ &= E(X^3) + 3\mu E(X^2) + 3\mu^2 E(X) - \mu^3 \\ &= E(X^3) - 3\mu(\mu^2 + \sigma^2) + 3\mu^3 - \mu^3 \\ &= E(X^3) - 3\mu\sigma^2 - \mu^3. \end{aligned}$$

Thus,

$$\mu_3 = E(X^3) = 3\mu\sigma^2 + \mu^3.$$

Therefore,

$$\hat{\mu}_3 = 3\bar{X}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^2\right) + \bar{x}^3 = \mu^3 + 3\mu\sigma^2.$$

Since the moment estimators of μ and σ^2 are identical to the maximum likelihood estimators of μ and σ^2, the ME and the MLE of μ_3 are identical.

Homework 9

Question. (6.5.1) If $(x_1, \ldots, x_n)$ is a sample from an $N(\mu, \sigma^2)$ distribution, where μ_0 is known and $\sigma^2 \in (0, \infty)$ is unknown, determine the Fisher information.

Solution. Let $\theta = \sigma^2$ for notational simplicity. Then the pdf is

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \text{exp}\bigg(- \frac{(x-\mu_0)^2}{2\theta}\bigg).$$

Its logarithm is

$$\log(f_\theta(x)) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{(x-\mu_0)^2}{\theta^3}.$$

The second-order partial derivative is

$$\frac{\partial^2 \log(f_\theta(x)}{\partial\theta^2} = \frac{1}{2\theta^2} - \frac{(x-\mu_0)^2}{\theta^3}.$$

It follows then that the Fisher information is

$$I(\theta) = -E\bigg(\frac{1}{2\theta^2} - \frac{(x-\mu_0)^2}{\theta^3}\bigg) = \frac{1}{2\theta^2}.$$

Recall that $\theta = \sigma^2$, and so by substitution we have

$$\frac{1}{2\theta^2} = \frac{1}{2\sigma^4}.$$

Recall by the theorem from the notes that we have then

$$nI(\theta) = \frac{n}{2\sigma^4}.$$

Question. (6.5.2) If $(x_1, \ldots, x_n)$ is a sample from a Gamma(α_0, θ) distribution, where α_0 is known and $\theta \in (0, \infty)$ is unknown, determine the Fisher information.

Solution. The PDF is

$$f_\theta(x) = \frac{\theta^{\alpha_0} x^{\alpha_0 - 1}}{\Gamma(\alpha_0)} e^{-\theta_0 x}.$$

Its logarithm is

$$log f_\theta(x) = -log(\Gamma(\alpha_0)) + \alpha_0 log(\theta) + (\alpha_0 - 1) log(x) - \theta x.$$

The second-order partial derivative is

$$\frac{\partial^2 log f_\theta(x)}{\partial\theta^2} = -\frac{\alpha_0}{\theta^2}.$$

Thus, the Fisher information is

$$I(\theta) = -E\bigg(\frac{\partial^2 log(f_\theta(x)}{\partial\theta^2}\bigg) = \frac{\alpha_0}{\theta^2}.$$

Question. (6.5.3) If $(x_1, \ldots, x_n)$ is a sample from a Pareto(α) distribution where $\alpha > 0$ is unknown, determine the Fisher information

Solution. The logarithm of it's PDF is

$$log(f_\alpha(x)) = log(\alpha(1+x)^{-(\alpha+1)}) = log(\alpha) - (\alpha+1) log(1+x)$$

for $0 < x < \infty$. Its second-order partial derivative is

$$\frac{\partial^2 log(f_\alpha(x)}{\partial\alpha^2} = -\frac{1}{\alpha^2}.$$

Thus, the Fisher information is

$$I(\alpha) = \frac{1}{\alpha^2}$$

and by the theorem from the notes we have

$$nI(\alpha) = \frac{n}{\alpha^2}.$$

Question. (6.5.4) Suppose the number of calls arriving at an answering service during a given hour of hte day is Poisson(λ), where $\lambda \in (0, \infty)$ is unknown. The number of calls actually recieved during this hour was recorded for 20 days, and the following data was obtained.

9	10	7	12	11	12	5	13	9	9
7	5	16	13	9	5	13	8	9	10

Construct an approximate 0.95-confidence interval for λ. Assess the hypothesis that this is a sample from a Poisson(11) distribution. If you are going to decide the hypothesis is false when the P-value is less than 0.05, then compute an approximate power for this procedure when $\lambda = 10$.

Solution. The logarithm of the PMF is

$$log(f_\lambda(x) = -log(x!) + xlog(\lambda) - \lambda.$$

Its second-order partial derivative is

$$\frac{\partial log(f_\lambda(x))}{\partial \lambda^2} = -\frac{x}{\lambda^2}.$$

The Fisher information is

$$E\left(\frac{\partial log(f_\lambda(x))}{\partial \lambda^2}\right) = \frac{E(X)}{\lambda^2} = \frac{1}{\lambda}.$$

Using $\hat{\theta} = \bar{x}$ for the MLE, we have

$$\sqrt{n}(\bar{x} - \lambda) \sim N(0, \lambda).$$

Based on the data, we have $n = 20$ and $\bar{x} = 9.65$. Thus, the 95% confidence interval for λ is

$$9.65 \pm 1.96\frac{\sqrt{9.65}}{\sqrt{20}} = [8.2885, 11.0115].$$

To test $H_0 : \lambda = 11$ against $H_1 : \lambda \neq 11$, we conclude H_0 since the 95% confidence interval contains 11. For power, we write the rejection region

$$C = \{\bar{x} < 11 - 1.96\sqrt{11/n} \text{ or } \bar{x} > 11 + 1.96\sqrt{11/n}\}$$

$$= \{\bar{x} < 9.546 \text{ or } \bar{x} > 12.454\}.$$

If $\lambda = 10$, then $\sqrt{n}(\bar{x} - 10) \sim N(0, 10)$ and the power is

$$P(\bar{x} \leq C|\lambda = 10) = P(\bar{x} < 9.546) + P(\bar{x} > 12.454)$$

$$\approx \Phi\left(\frac{9.546 - 10}{\sqrt{10/20}}\right) + \left(1 - \Phi\left(\frac{12.454 - 10}{\sqrt{10/20}}\right)\right)$$

$$= \Phi(-0.64) + [1 - \Phi(3.47)]$$

$$= 0.2613.$$

Question. (6.5.5) Suppose the lifelengths in hours of lightbulbs from a manufacturing process are known to be distributed Gamma($2, \theta$), where $\theta \in (0, \infty)$ is unknown. A random sample of 27 bulbs was taken and their lifelengths measured with the following data obtained.

336.87	2750.71	2199.44	292.99	1835.55	1385.36	2690.52
710.64	2162.01	1856.47	2225.68	3524.23	2618.51	361.68
979.54	2159.18	1908.94	1397.96	914.41	1548.48	1801.84
1016.16	1666.71	1196.42	1225.68	2422.53	753.24	

Determine an approximate 0.90-confidence interval for θ.

Solution. The logarithm of the PDF is

$$log(f_\theta(x) = log(\theta^2 x e^{-\theta x}) = 2log(\theta) + log(x) - \theta x.$$

Its second-order partial derivative is

$$\frac{\partial^2 log(f_\theta(x))}{\partial \theta^2} = -\frac{2}{\theta^2} \to I(\theta) = \frac{2}{\theta^2}.$$

Using the MLE of θ as $\hat{\theta} = 2/\bar{x}$, we obtain

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}\left(\frac{2}{\bar{x}} - \theta\right) \sim N(0, \theta^2/2).$$

Based on the data, we obtain $\bar{x} = 1627.47$ and $\hat{\theta} = 0.001229$. Thus the 90% confidence interval is

$$\hat{\theta} \pm 1.645\sqrt{\frac{\hat{\theta}^2}{2n}} = 0.001229 \pm 1.645\sqrt{\frac{0.001229^2}{54}} = [0.0009539, 0.001504].$$

Question. (6.5.7) Suppose that the incomes (measured in thousands of dollars) above $20K can be assumed to be Pareto(α), where $\alpha > 0$ is unknown for a particular population. A sample of 20 is taken from the population and the following data obtained.

21.265	20.857	21.090	20.047	20.019	32.509	21.622	20.093
20.109	23.182	21.199	20.035	20.084	20.038	22.054	20.190
20.488	20.456	20.066	20.302				

Solution. The PDF is

$$f(x) = \alpha(1 + x)^{-(\alpha+1)}.$$

Then,

$$E(X) = \int_0^\infty \alpha x(1 + x)^{-(\alpha+1)} dx$$
$$= \frac{1}{\alpha - 1}.$$

The loglikelihood function is

$$l(\alpha) = nlog(\alpha) - (\alpha + 1)\sum_{i=1}^{n} log(1 + x_i) \to l'(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} log(1 + x_i).$$

Thus, the MLE is

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log(1 + x_i)}.$$

Based on the data, we choose x_i as the values of income minus 20. We obtain $n = 20$ and $\hat{\alpha} = 1.786$. Thus, the estimate of the Fisher information is $I(\hat{\alpha}) = \frac{1}{\hat{\alpha}^2}$. The 95% confidence interval for α is

$$\hat{\alpha} \pm 1.96\sqrt{\frac{1}{nI(\hat{\alpha}))}} = 1.786 \pm 1.96\left(\frac{1.786}{\sqrt{20}}\right)$$
$$= [1.0033, 2.5687].$$

Using $\mu = E(X) = \frac{1}{\alpha - 1}$, we obtain $\alpha = 1.2$ if $x + 25 - 20 = 5$. Since 1.2 is inside the confidence interval, we accept $H_0 : \mu = 5$.

Remark. *It's also valid if you use the original data. It should prove the 95% confdience interval as [0.1823, 0.4666]. If mean is 25, then* $\alpha = 1.04$, *indicating that* H_0 *is rejected.*

Question. (6.5.8) Suppose that $(x_1, \ldots, x_n)$ is a sample from an Exponential(θ) distribution. Construct an approximate left-sided γ-confidence interval for θ.

Solution. The loglikelihood function is

$$l(\theta) = \sum_{i=1}^{n} log(\theta)e^{-\theta x_i} = nlog(\theta) - n\bar{x}\theta.$$

We obtain

$$l'(\theta) = \frac{n}{\theta} - n\bar{x} \to \hat{\theta} = \frac{1}{\bar{x}}.$$

Then

$$log(f_\theta(x)) = log(\theta) - x\theta \to \frac{\partial^2 log(f_\theta(x))}{\partial\theta^2} = -\frac{1}{\theta^2} - x \to I(\theta) = \frac{1}{\theta^2}.$$

Thus,

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \theta^2),$$

implying that an approximate left-sided γ confiene interval for θ is

$$(-\infty, \hat{\theta} + z_\gamma \hat{\theta}/\sqrt{n}].$$

Chapter 7

Homework 10

Question. (7.1.1) Suppose that $S = \{1, 2\}$, $\Omega = \{1, 2, 3\}$, and the class of probability distribution for the response s is given by the following table

	s=1	s=2
$f_1(s)$	1/2	1/2
$f_2(s)$	1/3	2/3
$f_3(s)$	3/4	1/4

If we use the prior $\pi(\theta)$ given by the table

	θ=1	θ=2	$\theta = 3$
$\pi(\theta)$	1/5	2/5	2/5

then determine the posterior distribution of θ for each possible sample of size 2.

Solution. The conditional PMF is

$$f_\theta(x) = \theta^{I(x=1)}(1-\theta)^{I(x=2)}.$$

The joint PMF is

$$f_\theta(x) = \theta^{I(x_1=1)+I(x_2=1)}(1-\theta)^{I(x_1=2)+I(x_2=2)}\pi(\theta_j),$$

where $\pi(\theta_j)$ is given by $\pi(1) = 0.2$, $\pi(2) = 0.4$, and $\pi(3) = 0.4$. The marginal PMF is

$$m(s) = \sum_{j=1}^{3} f_\theta(x)\pi(\theta_j).$$

The posterior PMF of θ is

$$q(\theta|x) = \frac{f_\theta(x)\pi(\theta)}{m(s)}.$$

We obtain the following table.

	θ=1	θ=2	$\theta = 3$
$\pi(\theta\|s = 1)$	3/16	1/4	9/16
$\pi(\theta\|s = 2)$	3/14	4/7	3/14

Question. (7.1.2) Suppose that we observe a sample $(x_1, \ldots, x_n)$ from the Bernoulli(θ) distribution with $\theta \in [0, 1]$ unknown. For the prior, we take π to be equal to a Beta(α, β) density. Find the mean and the variance of the posterior distribution.

Solution. To be explicit, note that

$$f_\theta(x) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$$

and

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \ x \in (0, 1).$$

To find the posterior distribution, we need to find $\int_\Omega f_\theta(x)\pi(\theta)d\theta$. Hence, we have

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \theta^{n\bar{x}+\alpha-1}(1-\theta)^{n+\beta-n\bar{x}-1}.$$

We can multiply the inside of the integral by $\frac{\Gamma(\alpha+n+\beta)}{\Gamma(n\bar{x}+\alpha)\Gamma(n+\beta-n\bar{x})}$ and the outside of the integral by it's inverse (since multiplied together this just gives one). Then we can note that the inside of the integral is Beta($n\bar{x}+\alpha, n(1-\bar{x})+\beta$) and so the integral over it's domain will give us with one, leaving us with

$$m(s) = \frac{\Gamma(n\bar{x}+\alpha)\Gamma(n(1-\bar{x})+\beta)}{\Gamma(\alpha+n+\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

Hence, we have

$$\pi(\theta|s) = \frac{f_\theta(x)\pi(\theta)}{m(s)},$$

and noting that we have a lot of cancellations, we get

$$\frac{\Gamma(\alpha+n+\beta)}{\Gamma(n\bar{x}+\alpha)\Gamma(n+\beta-n\bar{x})}\theta^{n\bar{x}+\alpha-1}(1-\theta)^{n-n\bar{x}+\beta-1},$$

which is a Beta$(n\bar{x}+\alpha, n(1-\bar{x})+\beta)$ distribution.
To find the mean and variance, we simply use the formulas from the tables from Chapter 1. Hence, we have

$$E(\theta|x) = \frac{n\bar{x}+\alpha}{n+\alpha+\beta}$$

and

$$Var(\theta|x) = \frac{(n\bar{x}+\alpha)[n(1-\bar{x})+\beta]}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}.$$

Question. (7.1.4) Suppose that $(x_1, \ldots, x_n)$ is a sample from a Poisson(λ) distribution with $\lambda \geq 0$ unknown. If we use the prior distribution for λ given by the Gamma(α, β) distribution, then determine the posterior distribution of λ.

Solution. For notational simplicity, we're going to assume $\lambda = \theta$ throughout. Then we have

$$f_\theta(x) = \frac{\theta^{n\bar{x}}}{\prod_{i=1}^n x_i!}e^{-n\theta}.$$

The prior PDF is

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}.$$

Using a similar strategy to the last problem, we have

$$m(s) = \frac{\beta^\alpha}{\Gamma(\alpha)\prod_{i=1}^n x_i!}\frac{\Gamma(n\bar{x}+\alpha)}{(n+\beta)^{n\bar{x}+\alpha}}$$

Combining this, we have that the posterior is

$$\pi(\theta|s) = \frac{(n+\beta)^{n\bar{x}+\alpha}}{\Gamma(n\bar{x}+\alpha)}\theta^{n\bar{x}+\alpha-1}e^{-(n+\beta)\theta},$$

which is the PDF of $\Gamma(n\bar{x}+\alpha, n+\beta)$.

Question. (7.1.5) Suppose that $(x_1, \ldots, x_n)$ is a sample from a Uniform$[0, \theta]$ distribution with $\theta > 0$ unknown. If the prior distribution of θ is given by Gamma(α, β), then determine the posterior density of θ.

Solution. The PDF of $X = (x_1, \ldots, x_n)$ is

$$f_\theta(x) = \frac{1}{\theta^n}I(0 \leq x_{(1)} \leq x_{(n)})I(x_{(1)} \leq x_{(n)} \leq \theta),$$

where $x_{(1)} = \min_{i \leq n}(x_i)$ and $x_{(n)} = \max_{i \leq n}(x_i)$. The prior PDF of θ is

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}.$$

Then we find that

$$m(s) = \int_0^\infty f_\theta(x)\pi(\theta)d\theta = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_{x_{(n)}}^\infty \theta^{\alpha-1-n}e^{-\beta\theta}d\theta.$$

It follows then that the posterior PDF of θ is

$$\pi(\theta|s) = \frac{\theta^{\alpha-1-n}e^{-\beta\theta}}{\int_{x_{(n)}}^\infty \theta^{\alpha-1-n}e^{-\beta\theta}d\theta}$$

for $\theta \geq x_{(n)}$.

Question. (7.1.9) Suppose you toss a coin and put a Uniform$[0.4, 0.6]$ prior on θ, the probability of getting a head on a single toss.
(a) If you toss the coin n times and obtain n heads, then determine the posterior density of θ.
(b) Suppose the true value of θ is, in fact, 0.99. Will the posterior distribution of θ ever put any probability mass around $\theta = 0.99$ for any sample of n?
(c) What do you conclude from part (b) about how you should choose a prior?

Solution. (a) Using $T = \sum_{i=1}^{n} \sim Bin(n, \theta)$, the PMF of T given θ is

$$f_\theta(T) = \binom{n}{T} \theta^T (1-\theta)^{n-T}, \; 0.4 \leq \theta \leq 0.6.$$

The prior PDF for θ is

$$\pi(\theta) = 5, \; 0.4 \leq \theta \leq 0.6.$$

The joint PMF-PDF of (T, θ) is

$$f_\theta(T)\pi(\theta) = \frac{5(n!)}{T!(n-T)!} \theta^T (1-\theta)^{n-T}.$$

The posterior PDF of θ is

$$\pi(\theta|T) = \frac{\theta^T (1-\theta)^{n-T}}{\int_{0.4}^{0.6} \theta^T (1-\theta)^{n-T} d\theta}, \; 0.4 \leq \theta \leq 0.6.$$

(b) Since the true value is outside $[0.4, 0.6]$, the posterior density of θ does not put any probability mass around 0.99.
(c) The prior density must be positive in the neighborhood of the true value of θ.

Question. (7.2.10) Suppose that $(x_1, \ldots, x_n)$ is a sample from the Exponential(λ) distribution, where $\lambda > 0$ is unknown and $\lambda \sim$ Gamma(α_0, β_0). Determine the mode of posterior distribution of λ. Also determine the posterior expectation and posterior variance of λ.

Solution. We have

$$f_\theta(x) = \theta^n e^{-n\bar{x}\theta}.$$

The prior for θ is

$$\pi(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0 - 1} e^{-\beta_0 \theta}.$$

We have that the joint PDF is

$$f_\theta(x)\pi(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{n+\alpha_0-1} e^{-\theta(n\bar{x}+\beta)}.$$

It follows then that

$$\pi(\theta|x) = \frac{(n\bar{x}+\beta)^{n+\alpha_0}}{\Gamma(n+\alpha_0)} \theta^{n+\alpha_0-1} e^{-\theta(n\bar{x}+\beta)},$$

which one can see is the PDF of a Gamma$(n+\alpha_0, n\bar{x}+\beta)$. Then

$$E(\theta|x) = \frac{n+\alpha_0}{n\bar{x}+\beta}$$

and

$$Var(\theta|x) = \frac{n+\alpha_0}{(n\bar{x}+\beta)^2}.$$

To compute the mode, we consider

$$log(\pi(\theta|s)) = log\left(\frac{(n\bar{x}+\beta)^{n+\alpha_0}}{\Gamma(n+\alpha_0)}\right) + (n+\alpha_0-1)log(\theta) - (n\bar{x}+\beta)\theta.$$

Let its first-order derivative be zero. We have

$$\frac{n+\alpha_0-1}{\theta} - (n\bar{x}+\beta) = 0 \rightarrow \theta_{mode} = \frac{n+\alpha_0-1}{n\bar{x}+\beta}.$$

Practice Midterm

Question. (a) Let $X_1, \ldots, X_n \sim N(\mu, 16)$ be a random sample. If $n = 100$ and $\bar{x} = 3.1$, then the 0,95-confidence interval for μ is (), the 0.99-confidence interval for μ is (). If we want the length of the 0.95-confidence interval to be less than or equal to 0.5, then we need n to be at least ().
(b) Suppose $X_1, \ldots, X_{100} \sim N(\mu, \sigma^2)$ be a random sample, with $\bar{x} = 4$ and $s^2 = 9$. Then, then 95% large sample confidence interval for μ is (), then 99% large sample confidence interval is (). If we want to test

$$H_0 : \mu = 2 \leftrightarrow H_1 : \mu \neq 2,$$

then H_0 is rejected if () at significance level $\alpha = 0.05$.
(c) Let $X \sim \text{Bin}(100, p)$. Suppose we observed $x = 32$. Then the 95% confidence interval for p is () and the 99% confidence interval for p is (). To make the 95% confidence interval less than or equal to 0.01, we need the sample size n at least ().

Solution. (a) Note that $z_{(1+0.95)/2} = 1.96$, and so we have $3.1 \pm 1.96\sqrt{\frac{16}{100}} = [2.316, 3.884]$. Note $z_{(1+0.99)/2} \approx 2.58$ (remember that we round up) and so we have $3.1 \pm 2.58\sqrt{\frac{16}{100}} = [2.068, 4.132]$. Finally, we have $2\delta = 0.5$, and so using the formula $n \geq \sigma^2\left(\frac{1.96}{\delta}\right)^2 \approx 984$ (remember again to round up).
(b) Since we're using the "large sample" confidence interval, then we still use the z-value instead of the t-value (since a large enough n for t is simply the z-value), and so we have $4 \pm 1.96\sqrt{\frac{9}{100}} = [3.412, 4.588]$. As a bonus, if we do not assume the large sample confidence interval, then we have $4 \pm 1.984\sqrt{\frac{9}{100}} = [3.4048, 4.5952]$. Next, we have $4 \pm 2.58\sqrt{\frac{9}{100}} = [3.226, 4.772]$. Finally, note that we reject the null hypothesis if $\left|\frac{\bar{x}-\mu}{\sqrt{s^2/100}}\right| > 100$. Plugging in the proper values, we get $\left|\frac{\bar{x}-2}{\sqrt{s^2/100}}\right| > 1.96$.
(c) Note that $\hat{p} = 32/100 = 0.32$, and $\hat{\sigma}^2 = \hat{p}(1-\hat{p}) = 0.32(1-0.32) = 0.2176$. So, we have $0.32 \pm 1.96\sqrt{\frac{0.2176}{100}} = [0.2286, 0.4114]$. Next, we have $0.32 \pm 2.58\sqrt{\frac{0.2176}{100}} = [0.19976, 0.4404]$. Finally, we have $2\delta = 0.01$, and so using the formula we have $\frac{1}{4}\left(\frac{1.96}{0.005}\right)^2 = 38416$.

Question. Assume $X_1 \sim N(2, 4)$, $X_2 \sim N(3, 6)$ and $X_3 \sim N(-3, 9)$.
(a) Compute $P(2X_1 + X_2 - X_3 < 16)$.
(b) Compute $P(|2X_1 - X_2 - X_3| < 7)$.
(c) Compute the density function of $2X_1 + X_3$.

Solution. (a) Let $Z \sim 2X_1 + X_2 - X_3$. Then we have $E(Z) = 2E(X_1) + E(X_2) - E(X_3) = 10$ and $V(Z) = 2^2V(X_1) + V(X_2) + V(X_3) = 31$. Hence, we have

$$P(Z < 16) = \Phi\left(\frac{16-10}{\sqrt{31}}\right) = 0.8594.$$

(b) Let $Z_2 \sim 2X_1 - X_2 - X_3$. Then this is equivalent to asking for $P(-7 < Z_2 < 7)$. Note that $E(Z_2) = 2E(X_1) - E(X_2) - E(X_3) = 4$ and $V(Z) = 2^2V(X_1) + V(X_2) + V(X_3) = 31$. Thus we have

$$P(-7 < Z < 7) = P(Z < 7) - P(Z < -7) = \Phi\left(\frac{7-4}{\sqrt{31}}\right) - \Phi\left(\frac{-7-4}{\sqrt{31}}\right) = 0.6809.$$

Remark. *This is Dr. Zhang's answer. I seem to not be getting the same thing using the same process. My answer is* ≈ 0.7275*. We had the same answer on (a), so it appears to be a typo.*

(c) Let $Z_3 \sim 2X_1 + X_3$. We have $E(Z_3) = 2E(X_1) + E(X_3) = 1$, and $V(Z_3) = 2^2V(X_1) + V(X_3) = 25$. Thus, $\mu = 1$ and $\sigma^2 = 25$, and so the density is

$$f(x) = \frac{1}{\sqrt{2\pi 5}}\exp\left(-\frac{(x-1)^2}{50}\right).$$

Question. Suppose $X_1, \ldots, X_n$ are iid with common PDF $f(x) = 2x/\theta^2$ for $0 \le x \le \theta$, with some $\theta > 0$. Let $X_{(n)} = \max(X_1, \ldots, X_n)$ be the sample maximum. Let $X_{(1)} = \min(X_1, \ldots, X_n)$ be the sample minimum.
(a) Compute the CDF and the PDF of $X_{(n)}$ and $X_{(1)}$.
(b) Compute $E(X_{(n)})$ and $V(X_{(n)})$.
(c) Compute $P(X_{(n)} > \theta - a)$ for some $a \in (0, \theta)$ and guess the behavior of the probability for large n.

Solution. (a) Here, we have to use a trick. Let F_n and F_1 be the CDF of $X_{(n)}$ and $X_{(1)}$ respectively, and likewise f_n and f_1 be the PDF of $X_{(n)}$ and $X_{(1)}$. Then we have

$$F_n(x) = F^n(x) = \left(\frac{x^2}{\theta^2}\right)^n = \frac{x^{2n}}{\theta^{2n}}$$

and

$$F_1(x) = 1 - [1 - F(x)]^n = 1 - \left(1 - \frac{x^2}{\theta^2}\right)^n.$$

Moreover, we have

$$f_n(x) = f_n'(x) = \frac{2nx^{2n-1}}{\theta^{2n}}$$

and

$$f_1(x) = f_1'(x) = n\left(1 - \frac{x^2}{\theta^2}\right)^{n-1}\left(-\frac{2x}{\theta^2}\right).$$

(b) By definition, we have

$$E(X_{(n)}) = \int_0^\theta x\frac{2nx^{2n-1}}{\theta^{2n}}dx = \frac{2n}{2n+1}\theta$$

and

$$E(X_{(n)}^2) = \int_0^\theta x^2\frac{2nx^{2n-1}}{\theta^{2n}}dx = \frac{2n}{2n+2}\theta^2.$$

Therefore, we have

$$V(X_{(n)}^2) = \left(\frac{2n}{2n+2} - \frac{(2n)^2}{(2n+1)^2}\right)\theta^2 = \frac{2n\theta^2}{(2n+2)(2n+1)^2}.$$

(c) We have

$$P(X_{(n)} > \theta - a) = 1 - F_n(\theta - a) = 1 - \frac{(\theta - a)^{2n}}{\theta^{2n}}.$$

The limit of the probability goes to 1 when $n \to \infty$.

Question. Let $\bar{X} = \sum_{i=1}^n X_i/n$.
(a) Suppose $X_1, \ldots, X_n$ are iid exponential distribution of mean θ, i.e., the density is $f(x) = e^{x/\theta}/\theta$ for $x > 0$, where $\theta > 0$ is a positive parameter. By the CLT, approximately compute $P(|\bar{X} - \theta| < 0.01\theta$ when n is 100, 10,000 and 1,000,000 respectively.
(b) Suppose X_i follows Poisson(2) distribution. Let μ and σ^2 be the common mean and variance of X_i. Then we have $\mu = \sigma^2 = 2$. By the CLT, approximately compute $P(\bar{X} < 2.2)$ when $n = 25, 31$. (Hint: use the continuity correction (0.5 shift)).

Solution. (a) First, we begin by rewriting the formula so that we have

$$P(|\bar{X} - \theta| < 0.01\theta) = P(-0.01\theta < \bar{X} - \theta < 0.01\theta).$$

Dividing both sides by $\sqrt{\theta^2/n}$, we have

$$\Phi\left(\frac{0.01\theta}{\sqrt{\theta^2/n}}\right) - \Phi\left(\frac{-0.01\theta}{\sqrt{\theta^2/n}}\right).$$

When $n = 100$, we have 0.0797; when $n = 10,000$, the value is 0.6827; and when $n = 1,000,000$, the value is 1.

(b) Using the definition for $n = 25$, we have

$$P(\bar{X} \leq 2.2) = P\left(\sum_{i=1}^{25} X_i \leq 55\right) = P\left(\sum_{i=1}^{25} X_i \leq 55.5\right)$$

$$= P(\bar{X} \leq 2.22) \approx \Phi\left(\frac{2.22 - 2}{\sqrt{2/25}}\right) = 0.7816.$$

Likewise, for $n = 31$, we have

$$P(\bar{X} \leq 2.2) = P\left(\sum_{i=1}^{31} X_i \leq 68.2\right) = P\left(\sum_{i=1}^{31} X_i \leq 68.5\right)$$

$$= P(\bar{X} \leq 2.2097) \approx \Phi\left(\frac{2.2097 - 2}{\sqrt{2/31}}\right) = 0.7954.$$

Question. Suppose $X_1, \dots, X_n$ are iid Gamma$(\alpha, 1)$. Let μ and σ^2 be the common mean and variance.
(a) Give the formulae of μ and σ^2 in terms of α.
(b) Show that $\bar{X}$ is an unbiased estimator of α.
(c) Give the MSE of $\bar{X}$ and show that it goes to 0 as $n \to \infty$.
(d) If the data is $1.39, 1.07, 0.72, 0.91, 2.56, 2.54, 0.91, 1.13, 4.72, 2.29$, compute the estimate value of α.

Solution. (a) This just follows by the tables. We have $\mu = \alpha$ and $\sigma^2 = \alpha$.
(b) Since $E(\bar{X}) = \alpha$, we have that it is an unbiased estimator of α.
(c) Since $V(\bar{X}) = (1/n^2)V(\sum_{i=1}^n X_i) = n\alpha/n^2 = \alpha/n$, we have

$$\text{MSE}(\bar{X}) = V(\bar{X}) = \alpha/n \to 0$$

as $n \to \infty$.
(d) Since $\mu = \alpha = E(X)$, we have then that it is 1.824.

Question. Suppose $X_1, \dots, X_n$ are iid Uniform$[0, \theta]$, i.e., the common density is $f(x) = 1/\theta$ if $0 \leq x \leq \theta$. Let $X_{(n)} = \max(X_1, \dots, X_n)$.
(a) Show that $X_{(n)}$ is a biased estimator of θ and compute the bias and the MSE.
(b) If one wants to modify a new estimator of θ based on $X_{(n)}$ by multiplying a constant c so that it is unbiased, what value of c do you suggest and find the MSE of the new estimator.
(c) Which estimator is better?

Solution. (a) The density of $X_{(n)}$ is $f_n(x) = nx^{n-1}/\theta^n$. It induces that $E(X_{(n)}) = n\theta/(n+1)$ and $V(X_{(n)}) = n\theta^2/[(n+2)(n+1)^2]$. The MSE is

$$\text{MSE}(X_{(n)}) = \left(\frac{n\theta}{(n+1)} - \theta\right)^2 + V(X_{(n)}) = \frac{2\theta^2}{(n+1)(n+2)}.$$

(b) It's clear that we select $c = (n+1)/n$. In this case

$$\text{MSE}(cX_{(n)}) = c^2 V(X_{(n)}) = \frac{\theta^2}{n(n+2)}.$$

(c) Since the MSE of $cX_{(n)}$ is smaller, then $\left(\frac{n+1}{n}\right)X_{(n)}$ is better.

Midterm

Question. (a) Suppose $E(X_1) = 1$, $E(X_2) = 2$, $V(X_1) = 1.44$, $V(X_2) = 2.56$, and $Cov(X_1, X_2) = 0.768$. Let $Y = 1.5X_1 - 1.2X_2$. Then $E(Y) = ()$, $V(Y) = ()$, $Corr(X_1, X_2) = ()$, and $Cov(X_1, Y) = ()$.
(b) Let (X, Y) be a bivariate random variable with joint PDF $f(x, y) = 3xy^2 + 2xy$, $0 < x, y < 1$. Then $P(X \leq 0.8, Y \leq 0.8) = ()$, $E(X) = ()$, $V(X) = ()$, and $Cov(X, Y) = ()$.
(c) Let $X_1, \ldots, X_{20}$ be iid $N(\mu, 16)$. Assume $\bar{x} = 4.32$. Then, the 95% confidence interval for μ is () and the 99% confidence interval for μ is (). If we want the 99% confidence interval to be shorter than 0.5, we need the sample size n to be at least ().
(d) Let $X \sim \text{Bin}(378, p)$. Suppose we observed $x = 249$. Then the 95% confidence interval for p is () and the 99% confidence interval for p is (). If we want to test $H_0 : p = 0.5$ against $H_1 : p \neq 0.5$ at 0.05 significance level, then we () the null hypothesis.
(e) Let $X_1, \ldots, X_n$ be iid with common mean $\mu = 1$ and common variance $\sigma^2 = 0.36$. Let $a = P(\bar{X} \leq 1.02)$. Using the CLT, if $n = 10^3$, then $a \approx ()$; if $n = 10^4$, then $a \approx ()$; and if $n = 10^5$, then $a \approx ()$.

Solution. (a) By the linearity of expectance, we have $E(Y) = 1.5E(X_1) - 1.2E(X_2) = -0.9$. Since X_1 and X_2 are not independent, we have $V(Y) = 1.5^2V(X_1) + 1.2^2V(X_2) + (2)(1.5)(-1.2)Cov(X_1, X_2) = 4.1616$. By the formula, we have $\text{Corr}(X_1, X_2) = \frac{0.768}{\sqrt{1.44*2.56}} = 0.4$. Finally, $\text{Cov}(X_1, Y) = \text{Cov}(X_1, 1.5X_1 - 1.2X_2) = (1.5)\text{Cov}(X_1, X_1) - (1.2)\text{Cov}(X_1, X_2) = (1.5)V(X_1) - 1.2\text{Cov}(X_1, X_2) = 1.2384$.
(b) Note that since $0 < x, y < 1$, we can assume x and y are independent. Then we have $P(X \leq 0.8, Y \leq 0.8) = \int_0^{0.8} \int_0^{0.8} 3xy^2 + 2xy dx dy = 0.36864$. We can then find the function $f_X(x)$ by integrating over all values of y; hence, we have $f_X(x) = \int_0^1 3xy^2 + 2xy dy = 2x$. Hence, we have $E(X) = \int_0^1 2x^2 dx = 2/3$. To find $V(X)$, we use the formula $V(X) = E(X^2) - E(X)^2$. Hence, we have $E(X^2) = \int_0^1 2x^3 dx = 1/2$, and so $V(X) = 1/2 - (2/3)^2 = 1/18$. One can quickly note that x and y are independent, and so $Cov(X, Y) = 0$.
(c) Note we have $n = 20$ and $\sigma^2 = 16$. Then by the formula we have that the 95% confidence interval is $4.32 \pm 1.96\sqrt{\frac{16}{20}} = [2.567, 6.073]$. Likewise, we have the 99% confidence interval as $4.32 \pm 2.58\sqrt{\frac{16}{20}} = [2.012, 6.628]$. Here, we have $2\delta = 0.5$, and so using the formula we have $16\left(\frac{2.58}{0.25}\right)^2 = 1705$.
(d) Since $x = 249$ and $n = 378$, we have $\hat{p} = \frac{249}{378} = 0.65873\ldots$. Then the 95% confidence interval for p is $0.6587 \pm 1.96\sqrt{\frac{(0.6587)(1-0.6587)}{378}} = [0.6109, 0.7065]$. Likewise, the 99% confidence interval for p is $0.6587 \pm 2.58\sqrt{\frac{(0.6587)(1-0.6587)}{378}} = [0.596, 0.722]$. If we want to test $H_0 : p = 0.5$, we have $\left|\frac{0.6587-0.5}{\sqrt{\frac{(0.6587)(1-0.6587)}{378}}}\right| = 0.00387$, and so we reject the null hypothesis. We could also just note that 0.5 is not in our 95% confidence interval.
(e) Let $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$. Then we have $P\big(Z \leq \frac{1.02-1}{\sqrt{0.36/n}}\big) = \Phi\left(\frac{1.02-1}{\sqrt{0.36/n}}\right)$. For $n = 10^3$, we have ≈ 0.8531, for $n = 10^4$, we have ≈ 0.9996, and finaly for $n = 10^5$ we have ≈ 1.

Question. Derive the maximum likelihood estimator (MLE) in the following problems. You need to display your entire approach (i.e., the maximum likelihood estimation).
(a) Let $X_1, \ldots, X_n$ be iid distributed of Poission(θ^2), $\theta > 0$. Derive the MLE of θ.
(b) Let $X_1, \ldots, X_n$ be iid distributed with $N(0, \theta)$, $\theta > 0$. Derive the MLE of θ.
(c) Let $X_1, \ldots, X_n$ be iid distributed with a common PDF $f_\theta(x) = \theta^2 xe^{-\theta x}$, $x, \theta > 0$. Derive the MLE of θ.
(d) Let $X_1, \ldots, X_n$ be iid distributed with a common PDF $f_\theta(x) = (\theta - 1)x^\theta$, $x \in (0, 1)$ and $\theta > -1$. Derive the MLE of θ.

Solution. (a) We have the PDF as

$$f_\theta(x_i) = \frac{e^{-\theta^2}\theta^{2x_i}}{x_i!}.$$

It follows then that the likelihood function is

$$L(\theta|s) = \prod_{i=1}^{n} \frac{e^{-\theta^2}\theta^{2x_i}}{x_i!} = \frac{e^{-n\theta^2}\theta^{2\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

and the loglikelihood function is

$$l(\theta|s) = -n\theta^2 + 2\sum_{i=1}^{n} x_i \log(\theta) + \log\bigg(\prod_{i=1}^{n} x_i!\bigg).$$

Taking the derivative, we have

$$l'(\theta|s) = -2n\theta + \frac{2\sum_{i=1}^{n} x_i}{\theta}$$

and setting it equal to 0 gives us

$$\hat{\theta} = \sqrt{\bar{x}}.$$

(b) First, let's note that the PDF is

$$f_\theta(x_i) = \frac{1}{\sqrt{2\pi}}\frac{1}{\theta}e^{-\frac{x_i^2}{2\theta}}.$$

The likelihood function is then

$$L(\theta|s) = \prod_{i=1}^{n}\left(\frac{1}{2\pi}\right)^{1/2}\left(\frac{1}{\theta}\right)^{1/2} e^{-\frac{x_i^2}{2\theta}}$$

$$= (2\pi)^{-n/2}(\theta)^{-n/2}e^{-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2}.$$

Hence, the loglikelihood function is then

$$l(\theta|s) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\theta) - \frac{1}{2\theta}\sum_{i=1}^{n} x_i^2$$

and, taking the derivative and setting it equal to zero, we get

$$l'(\theta|s) = -\frac{n}{2\theta} + \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} \to \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

(c) We first find the likelihood function. We have

$$L(\theta|s) = \prod_{i=1}^{n}\theta^2 x_i e^{-\theta x_i} = \theta^{2n}e^{-\theta\sum_{i=1}^{n} x_i}\prod_{i=1}^{n} x_i.$$

It follows then that the loglikelihood function is

$$l(\theta|s) = \sum_{i=1}^{n} x_i + 2nlog(\theta) - \theta\sum_{i=1}^{n} x_i.$$

Taking it's derivative and setting it equal to zero then gives us

$$l'(\theta|s) = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i \to \hat{\theta} = \frac{2}{\bar{x}}.$$

So, the MLE is $\hat{\theta} = 2/\bar{x}$.
(d) We find the likelihood function to be

$$L(\theta|s) = \prod_{i=1}^{n}(\theta-1)x_i^\theta = (\theta-1)^n\prod_{i=1}^{n} x_i^\theta.$$

The loglikelihood is then

$$l(\theta|s) = nlog(\theta-1) + \theta\sum_{i=1}^{n} log(x_i).$$

Taking the derivative and setting it equal to zero gives us

$$l'(\theta|s) = \frac{n}{\theta-1} + \sum_{i=1}^{n} log(x_i) \to \hat{\theta} = 1 - \frac{n}{\sum_{i=1}^{n} log(x_i)}.$$

Remark. *The PDF was originally $f_\theta(x) = (\theta+1)x^\theta$; he changed it on the exam to be $f_\theta(x) = (\theta-1)x^\theta$.*

Question. The following problems are related to the concepts of bias and mean square error (MSE).
(a) Let $X_1, \ldots, X_n$ be iid distributed of $N(\theta, 10)$. Compute the bias and the MSE of $\hat{\theta} = \hat{X}$.
(b) Let $X \sim \text{Bin}(n, \theta)$. Compute the bias and the MSE of $\hat{\theta} = X/n$.
(c) Let $X_1, \ldots, X_n$ be iid distributed with a common PDF $3x^2/\theta^3$, $0 \le x \le \theta$ for some $\theta > 0$. Compute the bias and the MSE of $\hat{\theta} = \max_{i \le n}(X_i)$.

Solution. (a) We have $E(\bar{X}) = \theta$, and $V(\hat{\theta}) = V(\sum_{i=1}^n X_i/n) = 1/n^2 V(\sum_{i=1}^n X_i) = n(10)/n^2 = 10/n$. Therefore,

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$$

and

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + \text{Bias}^2(\hat{\theta}) = V(\hat{\theta}) = 10/n.$$

(b) We have $E(X) = n\theta$ and $V(X) = n\theta(1-\theta)$, $E(\hat{\theta}) = \theta$ and $V(\hat{\theta}) = \theta(1-\theta)/n$. Thus,

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$$

and

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + \text{Bias}^2(\hat{\theta}) = V(\hat{\theta}) = \frac{\theta(1-\theta)}{n}.$$

(c) First, note that the CDF is x^3/θ^3 for some $\theta > 0$. Then we have

$$F(x) = P(\hat{\theta} \le x) = \prod_{i=1}^n P(X_i \le x) = \left(\frac{x^3}{\theta^3}\right)^n = \frac{x^{3n}}{\theta^{3n}}.$$

The PDF of $\hat{\theta}$ is

$$f(x) = F'(x) = \frac{3nx^{3n-1}}{\theta^{3n}}.$$

Then,

$$E(\hat{\theta}) = \int_0^\theta xf(x)dx = \frac{3n}{\theta^{3n}} \int_0^\theta x^{3n} dx = \frac{3n\theta}{3n+1}$$

and

$$E(\hat{\theta}^2) = \int_0^\theta x^2 f(x)dx = \frac{3n}{\theta^{3n}} \int_0^\theta x^{3n+1} dx = \frac{3n\theta^2}{(3n+2)(3n+1)^2.}$$

Thus,

$$V(\hat{\theta}) = E(\hat{\theta}^2) - E^2(\hat{\theta}) = \frac{3n\theta^2}{3n+2} - \left(\frac{3n\theta}{3n+1}\right)^2 = \frac{3n\theta^2}{(3n+2)(3n+1)^2}.$$

Thus,

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = \frac{\theta}{3n+1}$$

and

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + \text{Bias}^2(\hat{\theta}) = \frac{2\theta^2}{(3n+2)(3n+1)}.$$

Question. Let $X_1, \ldots, X_{25} \sim N(\mu, 16)$. Consider a hypothesis testing problem for

$$H_0 : \mu \ge 5 \leftrightarrow H_1 : \mu < 5.$$

Assume the rejection region is $C = \{\bar{X} \le 3.5\}$.
(a) Compute the Type I error probability at $\mu = 6$.
(b) Compute the Type II error probability at $\mu = 3$.
(c) Compute the significance level of the test.

Solution. (a) Note that $\bar{X} \sim N(\mu, 16/25) = N(\mu, 0.64)$. The Type I error probability at $\mu = 6$ is

$$P(\bar{X} \leq 3.5|\mu = 6) = \Phi\left(\frac{3.5 - 6}{\sqrt{0.64}}\right) = \Phi(-3.125) = 0.0006.$$

(b) The type II error probability at $\mu = 3$ is

$$P(\bar{X} > 3.5|\mu = 3) = 1 - \Phi\left(\frac{3.5 - 3}{\sqrt{0.64}}\right) = 1 - \Phi(0.625) = 0.2660.$$

(c) The significance level of the test is

$$P(\bar{X} \geq 3.5|\mu = 5) = \Phi\left(\frac{3.5 - 5}{\sqrt{0.64}}\right) = \Phi(-1.875) = 0.0304.$$

Practice Final

Question. Fill in the blanks.
(a) Let X_1 and X_2 be random variables with $E(X_1) = E(X_2) = 3$, $V(X_1) = V(X_2) = 5$, and $Corr(X_1, X_2) = 0.3$. Let $Y = 2X_1 + X_2$. Then $Cov(X_1, X_2) = ()$, $E(Y) = ()$, $V(Y) = ()$, and $Cov(X_2, Y) = ()$.
(b) Let $X \sim \text{Bin}(n, p)$. If $n = 100$ and $\hat{p} = X/n = 0.8$, then the 95% confidence interval for p is (). If we want the length of the 95% confidence interval for p to be less than or equal to 0.03, we need n at least ().
(c) Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$. Then, the maximum likelihood estimator of μ is (), the maximum likelihood estimator of σ^2 is (), and the maximum likelihood estimator of μ/σ is ().
(d) Let $X_1, \ldots, X_n$ be iid Uniform(θ). Then the maximum likelihood estimator of θ is (), the moment estimator of θ is (), and the mean square error (MSE) of the moment estimator is ().
(e) Let $X_1, \ldots, X_n$ be iid with common PDF $f_\theta(x) = \frac{1}{2}\theta^3 x^2 e^{-\theta x}$. Then, the MLE of θ is $\hat{\theta} = ()$ and the Fisher information is (). If n is large, the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is ().

Solution. (a) Recall that $Corr(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{V_1 V_2}}$. Hence, $Cov(X_1, X_2) = \sqrt{V_1 V_2} Corr(X_1, X_2)$. So, we have $5(0.3) = 1.5 = Cov(X_1, X_2)$. By linearity, we have $E(Y) = 2E(X_1) + E(X_2) = 9$. Since X_1 and X_2 are not independent, we have $V(Y) = 4V(X_1) + V(X_2) + (4)Cov(X_1, X_2) = 31$. Finally, we have $Cov(X_2, Y) = Cov(X_2, 2X_1 + X_2) = 2Cov(X_2, X_1) + V(X_2) = 8$.
(b) Since $\hat{p} = 0.8$, we have the 95% confidence interval is $0.8 \pm 1.96\sqrt{\frac{0.8(1-0.8)}{100}} = [0.7216, 0.8704]$. Using the formula, we have $n \geq \frac{1}{4}\left(\frac{1.96}{(0.03/2)}\right)^2 \to n = 4269$.
(c) We derived this in the notes. We have that the MLE of μ is $\bar{X}$. The MLE of σ^2 is then $\frac{1}{n}(\sum_{i=1}^n (X_i - \bar{X})^2$. Finally, we just combine these together to get $\frac{\bar{x}}{((1/n)(\sum_{i=1}^n (X_i - \bar{X})^2))^{1/2}}$
(d) This also comes from a homework problem. We have that the MLE of θ is $\max_{i \leq n}(X_i)$. Note that we have $m_1 = E(X)$, and $E(X) = \theta/2$, and so we have the moment estimator is $\theta = 2m_1 = 2\bar{X}$. Note that the bias is Bias $= E(\theta) - \theta = E(2\bar{X}) - 2\bar{X} = 0$, and so we have it is unbiased. Finally, $V(\theta) = V(2\bar{X}) = 4/n^2 V(\sum_{i=1}^n X_i) = 4n\theta^2/12n^2 = \theta^2/(3n)$.
(e) Note that the likelihood function is

$$L(\theta|s) = \prod_{i=1}^n \frac{1}{2}\theta^3 x_i^2 e^{-\theta x_i} = \frac{\theta^{3n}}{2^n} e^{-\theta \sum_{i=1}^n x_i} \prod_{i=1}^n x_i^2.$$

Then we have that the loglikelihood function is

$$l(\theta|s) = 3nlog(\theta) - nlog(2) - \theta \sum_{i=1}^n x_i + log\left(\prod_{i=1}^n x_i^2\right).$$

The derivative of the loglikelihood function is then

$$l'(\theta|s) = \frac{3n}{\theta} - \sum_{i=1}^n x_i \to \hat{\theta} = 3/\bar{x}.$$

Now, we need to find the Fisher information. We take the second derivative to get

$$l''(\theta|s) = -\frac{3n}{\theta^2}.$$

Thus, by definition, we have that the Fisher information is

$$E(-l''(\theta|x)) = \frac{3}{\theta^2}.$$

(Note that we omit the n, since what we actually derived was $nI(\theta)$.) Finally, recall that we have that

$$\sqrt{n}(\hat{\theta} - \theta) \to N(0, I^{-1}(\theta))$$

and so

$$\sqrt{n}(\hat{\theta} - \theta) \to N(0, \theta^2/3).$$

Question. Compute the following estimators of θ.
(a) Suppose $Y_1, \dots, Y_n$ are iid with common density $f(y) = \theta^2 y e^{-\theta y}$ for $0 \le y < \infty$ and $f(y) = 0$ for $y < 0$. Compute the maximum likelihood estimator of θ.
(b) Compute the moment estimator in (a).
(c) Suppose $Y_1, \dots, Y_n$ are iid with common density $f(y) = (\theta + 1)y^\theta$ for $0 < y < 1$ and $\theta > -1$. Find the moment estimator of θ.
(d) Compute the maximum likelihood estimator of θ for the problem given in (c).

Solution. (a) The likelihood function is

$$L(\theta|s) = \prod_{i=1}^{n} \theta^2 y e^{-\theta y} = \theta^{2n} e^{-\theta \sum_{i=1}^{n} y_i} \prod_{i=1}^{n} y_i.$$

It follows then that the loglikelihood function is

$$l(\theta|s) = 2nlog(\theta) - \theta \sum_{i=1}^{n} y_i + log(\prod_{i=1}^{n} y_i).$$

The derivative is then

$$l'(\theta|s) = \frac{2n}{\theta} - \sum_{i=1}^{n} y_i.$$

Thus, setting it equal to 0, we have

$$\hat{\theta} = \frac{2}{\bar{y}}.$$

(b) Note that

$$\mu = \int_0^\infty y f(y) dy = \int_0^\infty \theta^2 y^2 e^{-\theta y} dy = \frac{2}{\theta}.$$

The moment estimator is obtained then as followed

$$\bar{y} = \hat{\mu} = \frac{2}{\hat{\theta}} \to \hat{\theta} = \frac{2}{\bar{y}}.$$

(c) Note that

$$\mu = \int_0^1 y f(y) dy = \int_0^1 (\theta + 1) y^{\theta + 1} dy = \frac{\theta + 1}{\theta + 2}.$$

The moment estimator is then

$$\bar{y} = \hat{\mu} = \frac{\hat{\theta} + 1}{\hat{\theta} + 2} \to \hat{\theta} = \frac{2\bar{y} - 1}{1 - \bar{y}}.$$

(d) The likelihood function is

$$L(\theta|s) = \prod_{i=1}^{n} (\theta + 1) y_i^\theta = (\theta + 1)^n \prod_{i=1}^{n} y_i^\theta.$$

The loglikelihood function is

$$l(\theta|s) = nlog(\theta + 1) + \theta \sum_{i=1}^{n} log(y_i).$$

The derivative is

$$l'(\theta|s) = \frac{n}{\theta + 1} + \sum_{i=1}^{n} log(y_i) \to \hat{\theta} = -\frac{n}{\sum_{i=1}^{n} log(y_i)} - 1.$$

Question. A medical researcher wishes to determine if a pill has the undesirable side effect of reducing the blood pressure of the user. The study involves recording the initial blood pressures of 40 college-age women. Suppose the data are normally distributed. After they use the pill regularly for six months, their blood pressure are again recorder. The difference of the two measurement for each person was recorded. The sample mean $\bar{y} = 8.82$ and the sample variance $s^2 = 300.92$.
(a) Can you use the large-sample test for this problem? Why?
(b) Does the data indicate a significant reduction of blood pressure using a two-tailed test at level $\alpha = 0.01$.
(c) What is the p-value the test of part (a)?

Solution. (a) This is unimportant according to Dr. Zhang. However, the answer is that since $n = 40 > 30$, the sample size is large enough, and so we can use the large sample test.
(b) The test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. The statistic is

$$Z = \frac{8.82}{\sqrt{300.92/40}} = 3.216.$$

Since $|Z| = 3.216 > 2.58$, the null hypothesis is rejected and we conclude the reduction is significant.
(c) The p-value is

$$2[1 - \Phi(3.216)] = 0.0013.$$

Question. A market manager wishes to determine if lemon-scneted and almond-scented dishwashing liquids are equally likely by consumers. Out of a survey with 250 consumers interviewed, 145 prefer the lemon-scented and 105 prefer the almond-scented.
(a) Test whether the two liquids are equally likely at level 0.01.
(b) Compute the p-value of the test in (a).
(c) Based on the p-value in (a), which conclusion can you draw if $\alpha = 0.05$ and why?

Solution. (a) We need to test $H_0 : p = 0.5$ versus $H_1 : p \neq 0.5$. Since

$$\left| \frac{\hat{p} - 0.5}{\sqrt{(0.5)(1 - 0.5)/250}} \right| = \frac{0.08}{0.0316} = 2.54 < z_{0.005} = 2.58,$$

we accept H_0 and say there is not significant preference between those two liquids.
(b) The p-value is $2[1 - \Phi(2.54)] = 0.011$.
(c) Since $0.011 < 0.05$ the null hypothesis is rejected if $\alpha = 0.05$. In this case, we conclude they are significantly different.

Question. Let $X_1, \ldots, X_n \sim \text{Poisson}(\theta)$. The prior density of θ is

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha - 1} e^{-\beta x}, \beta > 0.$$

(a) Derive the marginal PMF of $X_1, \ldots, X_n$.
(b) Derive the posterior PDF of θ given $X_1, \ldots, X_n$.
(c) Under the square error loss $L(\delta, \theta) = (\delta - \theta)^2$, derive the Bayes estimator of θ.

Solution. (a) The conditional PMF of $X_1, \ldots, X_n$ given θ is

$$f_\theta(X_1, \ldots X_n) = \frac{\theta^{n\bar{X}} e^{-n\theta}}{\prod_{i=1}^n X_i!}.$$

The joint PMF-PDF of $X_1, \ldots, X_n$ and θ is

$$f_\theta(X_1, \ldots, X_n)\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha) \prod_{i=1}^n X_i!} \theta^{n\bar{X} + \alpha - 1} e^{-(n+\beta)\theta}.$$

Integrating θ out, we obtain the marginal PDF of $X_1, \ldots, X_n$ as

$$m(s) = \frac{\beta^\alpha}{\Gamma(\alpha) \prod_{i=1}^n X_i!} \frac{\Gamma(n\bar{X} + \alpha)}{(n + \beta)^{n\bar{X} + \alpha}}.$$

(b) The posterior PDF of θ is

$$\pi(\theta | s) = \frac{f_\theta(X_1, \ldots, X_n)\pi(\theta)}{m(s)} = \frac{(n + \beta)^{n\bar{X} + \alpha}}{\Gamma(n\bar{X} + \alpha)} \theta^{n\bar{X} + \alpha - 1} e^{-(n+\beta)\theta},$$

which is the PDF of $\Gamma(n\bar{X} + \alpha, n + \beta)$.
(c) The Bayesian estimator is

$$\hat{\theta}_{\text{Bayes}} = \int_0^\infty \theta \pi(\theta | X_1, \ldots, X_n) d\theta = \frac{n\bar{X} + \alpha}{n + \beta}.$$

This is also the mean of $\Gamma(n\bar{X} + \alpha, n + \beta)$.