2022 Probability and Statistics Final Solutions

1. Show that the variance of a binomial distribution $b(n, p)$ is $\sigma^2 = npq = np(1-p)$. (You may use any other information about the distribution from the distribution table.)

Solution

To compute the variance of $X \sim b(1,p)$ we compute the second moment $E(X^2) = p(1^2) = p$, and so the variance is $p - E(X)^2 = p(1-p) = pq$. For $Y \sim b(n,p)$ we use that $Y = \sum X_i$ is the sum of n independent $b(1,p)$ RVs so the variance is $Var(Y) = \sum Var(X_i) = npq$.

Alternately we could use the moment generating function $M_Y(t) = (pe^t + q)^n$. Taking its first two derivatives

$$M_Y'(t) = n(pe^t + q)^{n-1}pe^t$$

$$M_Y''(t) = n(pe^t + q)^{n-1}pe^t + n(n-1)(pe^t + q)^{n-2}(pe^t)^2$$

and evaluating them at 0 we get

$$E(X) = np \qquad E(X^2) = M_Y''(0) = np + n(n-1)p^2 = np + (np)^2 - np^2$$

so $\sigma^2 = np + (np)^2 - np^2 - (np)^2 = np - np^2 = npq$, as needed.

2. A grocer sells, on average, 5 pumpkins a week, (according to a poisson distribution). How many should he stock so that the chance of running out by the end of the week is less than .01. (We may assume that pumpkins will sell according to a poisson distribution.)

Solution

Where X is a poisson distribution with mean 5 we want the value x such that $P(X > x) < .01$ Using tables we find that $P(X > 10) = .014$ and $P(X > 11) = .005$. He should stock 11.

3. In a sample of 60 randomly chosen people from a population, 15 are left handed. Give, (using the central limit theorem) an approximate .95 confidence interval for percentage of the population that is left handed.

Solution

I asked for a confidence interval for the 'percentage of the population' but most people found a confidence interval for the probability p that a given person is left-handed. This is really what I meant, so we find this first, then we multiply it by 100 to get the confidence interval for the percentage of the population.

As we are using the CLT, many people knew that the confidence interval for p should be

$$[\overline{X} - z_{.975}S/\sqrt{n}, \overline{X} + z_{.975}S/\sqrt{n}].$$

Most people got that our estimate of p is $\overline{X} = 15/60 = .25$, that the sample size is $n = 60$, and that the z-value is $z_{.975} = 1.96$. What many people had trouble with was S.

Our sample is a random sample $\mathbf{X} = (X_1, \ldots, X_{60})$ from a distribution X. The sum $Y = \sum X_i = 15$ is the number of left handed people. So what is X? It has two possible values. It can only be a Bernoulli random variable $X \sim b(1, p)$ where p is the probablility that a random person is left handed. So its variance is $\sigma^2 = pq$, giving us a sample variance of $S^2 = (.25)(.75) = .1875$. Putting these into the above, the confidence interval for p is

$$[.25 - 1.96(\sqrt{.1875}/\sqrt{60}), .25 + 1.96(\sqrt{.1875}/\sqrt{60})] \approx [.23, .37].$$

(The approximate confindence interval for the percentage of the population that is left handed is $[23, 37]$)

4. Let Y_i be the i^{th} order statistic of a random sample $\mathbf{X} = (X_1, \ldots, X_n)$ of a continuous random variable X.

   (a) Find the probability $P(Y_n < Y_1)$.

   (b) Find the probabiliy $P(X_1 \leq Y_1)$.

   (c) Find the probability $P(\xi_{.25} < Y_i < \xi_{.75})$.

Solution

The first two parts of this question were supposed to be free marks. Maybe it is because they are so easy that almost nobody got them.

(i) 0. The order statistics are ordered $Y_1 \leq Y_2 \leq \cdots \leq Y_n$. Y_n can equal Y_1 but it cannot be less than Y_1.

(ii) $1/n$. The event $X_1 \leq Y_1$ occurs when X_1 is the smallest of the X_i. As the X_i are iid, the all have the same probablility of being the smallest. So the probability that X_1 is the smallest is $1/n$. (As the X_i are continuous, the probability that two are the same is 0.)

(iii) This is the probability that at least i of the X_i are less than $\xi_{.75}$, minus the probability that at least i are less than $\xi_{.25}$. (Again, as they are continuous we can replace $<$ with $\leq$ without changing probabilities.) By definition we have $P(X < \xi_p) = p$, so this is

$$\sum_{\alpha=i}^{n} \binom{n}{\alpha} (.75)^{\alpha} (.25)^{n-\alpha} - \sum_{\alpha=i}^{n} \binom{n}{\alpha} (.25)^{\alpha} (.75)^{n-\alpha}.$$

5. Let $X_1, X_2, \ldots, X_{20}$ be a random sample of normal distribution X with mean μ. For the hypotheses $H_0 : \mu = 10$ and $H_1 : \mu \neq 10$, the critical region (for the parameter $\overline{X}$) is $C = (12, \infty)$. We get a sample realisation with mean $\overline{x} = 11.5$ and $S^2 = 5$. What are the significance of the test, and the power of the test for the value $\mu = 13$?

Solution

The significance of the test; or probability of a false positive; is $Prob(\overline{x} > 12 \mid \mu = 10)$. Using our sample variance of $S^2 = 5$ to estimate the variance σ^2 of X, our significance is

$$\begin{aligned} \text{Prob}(\overline{X} > 12 \mid \mu = 10) &= \text{Prob}(T_{19} > \frac{12 - \mu}{S/\sqrt{n}}) \\ &= \text{Prob}(T_{19} > \frac{12 - 10}{\sqrt{5/20}}) = P(T_{19} > 4). \end{aligned}$$

Using the t-table, $t_{19,0.999} = 3.579$. As 4 is bigger than this, this significance is less than .001.

The power $\gamma(13)$ of the test for a value $\mu = 13$ is

$$\begin{aligned} \text{Prob}(\overline{X} > 12 \mid \mu = 13) &= \text{Prob}(T_{19} > \frac{12 - 13}{\sqrt{5/20}}) \\ &= \text{Prob}(T_{19} > -2) = \text{Prob}(T_{19} < 2) \approx .97. \end{aligned}$$

6. Each of 50 golfers hit a golf ball of brand X and a golf ball of brand Y. For each i, let X_i and Y_i be the distance that golfer i hit the respective

balls, and let $W_i = X_i - Y_i$. Where μ is the mean of the distribution of W, test the hypotheses

$$H_0 : \mu = 0 \qquad \text{vs.} \qquad H_1 : \mu > 0.$$

What is the p-value of the test for a realised sample mean $\overline{w} = 2.07$ and sample variance $S^2 = 75$?

Solution

The sample is large enough that we can use the CLT. The significance (p-value) is $\alpha = P(Z > \frac{2.07-0}{\sqrt{75/50}}) = P(Z > 2.07/\sqrt{1.5}) \approx P(Z > 2.07/1.2) \approx P(Z > 1.7)$. Checking the z-tables, this is about .0446.

Bernoulli $X \sim b(1,p)$

$p_X(x)$	$\begin{cases} p & \text{if } x = 1 \\ 1-p & \text{otherwise.} \end{cases}$
μ	p
σ^2	pq
$M_X(t)$	$pe^t + q$

Binomial $X \sim b(n,p)$

$p_X(x)$	$\binom{n}{x} p^x (1-p)^{n-x}$
μ	np
σ^2	npq
$M_X(t)$	$(pe^t + q)^n$

Hypergeometric

$p_X(x)$	$\frac{\binom{N-D}{n-x}\binom{D}{x}}{\binom{N}{n}}$
μ	$n\frac{D}{N}$
σ^2	$n\frac{D}{N}\frac{N-D}{N}\frac{N-n}{N-1}$

Poisson $X \sim \text{pois}(\mu)$

$p_X(x)$	$\frac{e^{-\mu}\mu^x}{x!}$
μ	μ
σ^2	μ
$M_X(t)$	$e^{\mu(e^t-1)}$

Gamma $X \sim \Gamma(\alpha,\beta)$

$f_X(x)$	$\frac{x^{\alpha-1}e^{-(x/\beta)}}{\Gamma(\alpha)\beta^\alpha}$
μ	$\alpha\beta$
σ^2	$\alpha\beta^2$
$M_X(t)$	$(1-\beta t)^{-\alpha}$

Exponential $X \sim \Gamma(1,\mu)$

$f_X(x)$	$e^{-x/\mu}/\mu$
μ	μ
σ^2	μ^2
$M_X(t)$	$(1-\mu t)^{-1}$

Normal $X \sim N(\mu,\sigma^2)$

$f_X(x)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
$M_X(t)$	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$

Table I
Poisson Distribution

The following table presents selected Poisson distributions. The probabilities tabled are

$$P(X \leq x) = \sum_{w=0}^{x} e^{-m} \frac{m^w}{w!},$$

for the values of m selected.

x	$m = E(X)$ 0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
0	0.607	0.368	0.223	0.135	0.050	0.018	0.007	0.002	0.001	0.000	0.000	0.000
1	0.910	0.736	0.558	0.406	0.199	0.092	0.040	0.017	0.007	0.003	0.001	0.000
2	0.986	0.920	0.809	0.677	0.423	0.238	0.125	0.062	0.030	0.014	0.006	0.003
3	0.998	0.981	0.934	0.857	0.647	0.433	0.265	0.151	0.082	0.042	0.021	0.010
4	1.000	0.996	0.981	0.947	0.815	0.629	0.440	0.285	0.173	0.100	0.055	0.029
5	1.000	0.999	0.996	0.983	0.916	0.785	0.616	0.446	0.301	0.191	0.116	0.067
6	1.000	1.000	0.999	0.995	0.966	0.889	0.762	0.606	0.450	0.313	0.207	0.130
7	1.000	1.000	1.000	0.999	0.988	0.949	0.867	0.744	0.599	0.453	0.324	0.220
8	1.000	1.000	1.000	1.000	0.996	0.979	0.932	0.847	0.729	0.593	0.456	0.333
9	1.000	1.000	1.000	1.000	0.999	0.992	0.968	0.916	0.830	0.717	0.587	0.458
10	1.000	1.000	1.000	1.000	1.000	0.997	0.986	0.957	0.901	0.816	0.706	0.583
11	1.000	1.000	1.000	1.000	1.000	0.999	0.995	0.980	0.947	0.888	0.803	0.697
12	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.991	0.973	0.936	0.876	0.792
13	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.987	0.966	0.926	0.864
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.983	0.959	0.917
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.992	0.978	0.951
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.989	0.973
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.986
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.993
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table II
Chi-square Distribution

The following table presents selected quantiles of chi-square distribution; i.e, the values x such that

$$P(X \leq x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2}\, dw,$$

for selected degrees of freedom r.

r	$P(X \leq x)$ 0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Table III
Normal Distribution

The following table presents the standard normal distribution. The probabilities tabled are

$$P(X \leq x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw.$$

Note that only the probabilities for $x \geq 0$ are tabled. To obtain the probabilities for $x < 0$, use the identity $\Phi(-x) = 1 - \Phi(x)$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Table IV
t-Distribution

The following table presents selected quantiles of the t-distribution; i.e, the values x such that

$$P(X \leq x) = \int_{-\infty}^{x} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1 + w^2/r)^{(r+1)/2}}\, dw$$

for selected degrees of freedom r. The last row gives the standard normal quantiles.

r	$P(X \leq x)$ 0.900	0.950	0.975	0.990	0.995	0.999
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
∞	1.282	1.645	1.960	2.326	2.576	3.090