Homework 1 Solutions

ESE 402/542

Due September 27, 2020 at 11:59pm

Type or scan your answers as a single PDF file and submit on Canvas.

Problem 1. Counting review.

(a) The game of Mastermind starts in the following way: One player selects four pegs, each peg having six possible colors, and places them in a line. The second player then tries to guess the sequence of colors. What is the probability of guessing correctly?

Solution. There are 6^4 ways for player 2 to guess the four pegs. Their probability of guessing correctly is $\frac{1}{6^4}$.

(b) A child has six blocks, three of which are red and three of which are green. How many patterns can she make by placing them all in a line? If she is given three white blocks, how many total patterns can she make by placing all nine blocks in a line?

Solution. First, we have 6 slots. We choose 3 slots for our red blocks, and so our green blocks are determined already. We get $\binom{6}{3} = \frac{6!}{3!3!} = 20$ ways.

Next, we are given 3 white blocks. We choose 3 of the 9 slots for red, 3 or the remaining 6 for green, and so the white blocks go in the remaining slots. We get $\binom{9}{3}\binom{6}{3} = \frac{9!}{3!3!3!} = 1680$ ways.

Problem 2. A random variable X is distributed according to $Gamma(\alpha, \beta)$. Recall that the pdf of a gamma random variable is given to be $f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$. Also recall that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$.

$\int t^{\alpha+1} e^{-t} dt = \Gamma(\alpha+2)$

(a) Show that $\mathbf{E}[X] = \alpha\beta$

Solution. We set up the expectation and make a u substitution of $u = x/\beta$ that results in a gamma function.

$$\begin{aligned}
\mathbf{E}[X] &= \int_0^\infty x f(x) dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^\alpha e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \beta^\alpha x^\alpha e^{-u} \beta du \\
&= \frac{\beta}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-u} du \quad = \Gamma(\alpha+1) \\
&= \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} \\
&= \alpha\beta
\end{aligned}$$

(b) Show that $\mathbf{Var}[X] = \alpha\beta^2$

Solution. We write variance as $\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$ and find find $\mathbf{E}[X^2]$.

$$\begin{aligned}
\mathbf{E}[X^2] &= \int_0^\infty x^2 f(x)dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+1}e^{-x/\beta}dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty u^{\alpha+1}\beta^{\alpha+1}e^{-u}\beta du \\
&= \frac{\beta^2}{\Gamma(\alpha)}\int_0^\infty u^{\alpha+1}e^{-u}du \\
&= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} \\
&= \beta^2(\alpha+1)\alpha
\end{aligned}$$

So we compute the variance as follows.

$$\mathbf{Var}[X] = \beta^2(\alpha+1)\alpha - (\beta\alpha)^2 = \beta^2\alpha$$

(c) Find α and β such that X is a chi squared random variable with p degrees of freedom.

Solution. $\alpha = \frac{p}{2}$ and $\beta = 2$. (We can directly compare the PDF of Gamma and chi squared distributions to get this value)

Problem 3. Find the moment-generating function of a random variable $X \sim Bernoulli(p)$, and use the MGF you computed to find its mean, variance, and third moment $\mathbf{E}[X^3]$. Recall that pmf of Bernoulli distribution is $f(x) = p^x(1-p)^{1-x}$.

Solution. The moment generating function, by definition, would be

$$M_x(t) = E[e^{tx}] = e^{t*1}P(x=1) + e^{t*0}P(x=0)$$

$$= pe^t + 1 - p$$

So $E[X] = M'(0) = pe^0 = p$,
$Var[X] = E[X^2] - (E[X])^2 = M''(0) - p^2 = pe^0 - p^2 = p - p^2$
$E[X^3] = M'''(0) = pe^0 = p$

Problem 4. Suppose that a company ships packages that are variable in weight, with an average weight of 30 lb and a standard deviation of 15. Assuming that the packages come from a large number of different customers so that it is reasonable to model their weights as independent random variables, find the probability that 100 packages will have a total weight exceeding 3500 lb.

Solution. Each package i $(0 \leq i \leq 100)$ can be modeled with a normally distributed random variable,

$$X_i \sim \mathcal{N}(\mu = 30, \sigma = 15)$$

Then, the sum of these 100 packages can be modeled by a normal r.v. Y with

$$\mu_Y = \sum_{i=0}^{100} \mu = 3000$$

$$\sigma_Y^2 = \sum_{i=0}^{100} \sigma^2 = 22500$$

because X_is are assumed to be independent.
The probability that 100 packages will have a total weight exceeding 1700 is

$$\begin{aligned} P(Y > 3500) &= \Phi(\frac{3500 - 3000}{150}) \\ &= 4.29 * 10^{-4} \end{aligned}$$

Answer= 4.29*10^{-4}

Problem 5. A particular area contains 8,000 condominium units. In a survey of the occupants with sample size 100, 12% of the respondents said they planned to sell their condos within the next year;

(a) Compute the 95% confidence interval for the estimated probability of people planning to sell. (Hint: The variance of the Bernoulli distribution with parameter p is given by $p(1-p)$).

(b) Suppose that another survey is done of another condominium project of 12,000 units. The sample size is 200, and the proportion planning to sell in this sample is .18. What is the standard error of this estimate? Give a 90% confidence interval.

(c) Suppose we use the notation $\hat{p}_1 = .12$ and $\hat{p}_2 = .18$ to refer to the proportions in the two samples. Let $\hat{d} = \hat{p}_1 - \hat{p}_2$ be an estimate of the difference, d, of the two population proportions p_1 and p_2. Using the fact that $\hat{p}_1$ and $\hat{p}_2$ are independent random variables, calculate the variance and standard error of $\hat{d}$.

(d) Because $\hat{p}_1$ and $\hat{p}_2$ are approximately normally distributed, so is $\hat{d}$. Use this fact to construct 99%, 95%, and 90% confidence intervals for d. Is there clear evidence that p_1 is really different from p_2?

Solution

(a) The CI can be given as

$$= \hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \hat{p} \pm 1.96 \times \sqrt{\frac{0.12 \times 0.88}{n}}$$

$$= (0.056, 0.184)$$

(b) Similarly, The CI is

$$= \hat{p} \pm z(0.05) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$= \hat{p} \pm z(0.05) \times \sqrt{\frac{0.18 \times 0.82}{n}}$$
$$= (0.136, 0.224)$$

The Std error is $\sqrt{\frac{0.18 \times 0.82}{200}} = 0.027$

(c) By independence of p_1 and p_2,

$$Var(\hat{d}) = Var(\hat{p_1}) + Var(\hat{p_2})$$
$$= \frac{0.12 \times 0.88}{100} + \frac{0.18 \times 0.82}{200} = 1.7827 \times 10^{-3}$$

Std error $= \sqrt{Var(\hat{d})} = 0.04$

$d = -0.06$

$\alpha = 0.01$, $\frac{\alpha}{2} = 0.005$, $z = 2.576$

$-0.06 \pm 1.96 \times 0.04$

$\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $z = 1.96$

$\alpha = 0.1$ $\frac{\alpha}{2} = 0.05$ $z = 1.645$

(d) The CI can be calculated with $\hat{d} \pm z(\frac{\alpha}{2}) * s_{\hat{d}}$
99% (-0.1697,0.0497)
95% (-0.1433,0.0233)
90% (-0.1239, 0.0099)
We can see that zero is in all three CIs, thus $\hat{p_1}$ is not really different from $\hat{p_2}$

Problem 6. Let $\hat{\theta}$ be an estimator for a scalar parameter θ of a parametric model. The mean squared error of θ is define as

$$\mathrm{MSE}(\hat{\theta}, \theta) = \mathbb{E}[|\hat{\theta} - \theta|^2]. \tag{1}$$

Show that

$$\mathrm{MSE}(\hat{\theta}, \theta) = (\mathrm{bias}(\hat{\theta}, \theta))^2 + \mathrm{Var}(\hat{\theta}) \tag{2}$$

where bias$(\hat{\theta}, \theta)$ is the the difference between the average of the estimate and the true parameter; i.e., $\mathbb{E}[\hat{\theta}] - \theta$.

Solution.

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}, \theta) &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\
&= \mathbb{E}[(\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2] \\
&= \mathbb{E}[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 + 2((\hat{\theta} - \mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta}) - \theta)) + (\mathbb{E}(\hat{\theta}) - \theta)^2] \\
&= \mathbb{E}[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2] + 2(\mathbb{E}(\hat{\theta}) - \theta)\mathbb{E}[\hat{\theta} - \mathbb{E}(\hat{\theta})] + (\mathbb{E}(\hat{\theta}) - \theta)^2 \\
&= \mathrm{Var}(\hat{\theta}) + (\mathrm{bias}(\hat{\theta}, \theta))^2
\end{aligned}$$

Or any reasonable derivation.

Problem 7. There are $p = 25\%$ of total population that will vote for candidate A in a coming election. In a survey of sample size to be $n = 100$, $\hat{p}$ of interviewee said they would vote for A.

(a) Find δ such that $P(|\hat{p} - p| \geq \delta) = 0.025$.

(b) If, in the sample, $\hat{p} = 0.25$, will the 95% confidence interval for p contain the true value of p?

Solution

(a) Std error $= \sqrt{\frac{0.25(1-0.25)}{100}} = 0.043$
Then using CLT, we have

$$P(|z| \leq \frac{\delta}{0.043}) = 0.975$$
$$P(z \leq \frac{\delta}{0.043}) = 0.9875$$
$$\frac{\delta}{0.043} = \Phi^{-1}(0.09875)$$
$$\delta = 9.63 \times 10^{-2}$$

(b) The CI is given by

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$= 0.25 \pm 1.96 \sqrt{\frac{0.25(1-0.25)}{100}}$$
$$= 0.25 \pm 0.08487$$
$$= (0.1651, 0.3349)$$

We can see that p = 0.25 is in the interval. Alternatively, we can reason that 0.25 will always be in a CI centered at 0.25 itself, without doing the calculation