X_400004 - Statistics
Solutions to the Mock Final Exam

9 December 2020

Below are answers to the exam questions. Some of these are slightly abbreviated, while others include extra comments. These are for your reference only but should inform the level of detail that is expected from your answers in the exam. Also keep in mind that there might be different ways to approach each question. If you find typos and or omissions, please report them to the lecturer so they can be corrected.

Prob.I: Consider a random sample $X_1, \dots, X_n$ of size n from the $\text{Ber}(p)$ distribution, where $p \in [0, 1]$ and suppose that you put a $\text{Beta}(\alpha, \beta)$ prior on p. This problem concerns the choice of the hyper-parameters α and β.

A few facts that you *may* need to know in order to solve this questions are:

- If $X \sim \text{Ber}(p)$, then the probability mass function of X is $f(x) = p^x(1-p)^{1-x}$;
- If $Y \sim \text{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$, then the probability density function of Y is $f(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}$;
- If $Y \sim \text{Beta}(\alpha, \beta)$, then $\mathbb{E}Y = \alpha/(\alpha+\beta)$, and $\mathbb{V}Y = \alpha\beta/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$;

(a) Compute the posterior distribution of p. **Solution: The posterior distribution on p is proportional to the likelihood times the prior. The likelihood is**

$$L(p) = p^{X_1}(1-p)^{1-X_1} \times \cdots \times p^{X_n}(1-p)^{1-X_n} = p^{n\bar{X}}(1-p)^{n-n\bar{X}}.$$

The prior density is $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$. Combining the two we get that the posterior density satisfies

$$\pi(p \mid X_1, \dots, X_n) \propto L(p)\pi(p) \propto p^{n\bar{X}}(1-p)^{n-n\bar{X}}p^{\alpha-1}(1-p)^{\beta-1} = p^{n\bar{X}+\alpha-1}(1-p)^{n-n\bar{X}+\beta-1}.$$

We identify this as being proportional to the density of a Beta$(n\bar{X}+\alpha, n-n\bar{X}+\beta)$ distribution. We conclude that the posterior is Beta$(n\bar{X}+\alpha, n-n\bar{X}+\beta)$.

(b) Compute the expectation of the posterior distribution and call is $\hat{p}$; this is your estimator of p. **Solution: Using the expression for the expectation of a Beta distribution, we have**

$$\hat{p} = \frac{n\bar{X}+\alpha}{n\bar{X}+\alpha+n-n\bar{X}+\beta} = \frac{n\bar{X}+\alpha}{n+\alpha+\beta}.$$

(c) Compute the bias and the variance of $\hat{p}$. (Note that these will depend on α and β.) **Solution: The expectation and variance of $\bar{X}$ are respectively p and $p(1-p)/n$, so**

$$Bias_{\alpha,\beta}(p) = \mathbb{E}\hat{p} - p = \frac{n\mathbb{E}\bar{X} + \alpha}{n+\alpha+\beta} - p = \frac{np + \alpha - (n+\alpha+\beta)p}{n+\alpha+\beta} = \frac{\alpha(1-p) - \beta p}{n+\alpha+\beta}.$$

The variance of $\hat{p}$ is

$$Var_{\alpha,\beta}(p) = \mathbb{V}\hat{p} = \frac{\mathbb{V}(n\bar{X}+\alpha)}{(n+\alpha+\beta)^2} = \frac{n^2\mathbb{V}(\bar{X})}{(n+\alpha+\beta)^2} = \frac{n^2 p(1-p)/n}{(n+\alpha+\beta)^2} = \frac{np(1-p)}{(n+\alpha+\beta)^2}.$$

(d) Is it possible to pick hyperparameters α, β (not depending on p) so that the resulting $\hat{p}$ is unbiased? If so, what is the resulting estimator? **Solution: For any α, setting $\beta = \alpha(1-p)/p$ leads to the respective $\hat{p}$ being unbiased. The only choice of this type that does not depend of p is to set $\alpha = \beta = 0$. Although technically this is not allowed for the prior, the resulting posterior is well defined when $\alpha = \beta = 0$ and has expectation $\bar{X}$.**

(e) The estimator $\bar{X}$ is the Maximum Likelihood estimator and as such has the smallest variance among any estimator of p. Is it possible to pick α, β so that the corresponding $\hat{p}$ has a **smaller Mean Squared Error** than the Maximum Likelihood estimator? (Justify your answer.) **Solution: As we just saw, setting $\alpha = \beta = 0$ leads to the posterior expectation being the Maximum Likelihood estimator which is unbiased and has the smallest variance of any estimator of p. The bias-variance decomposition then tells us that the Maximum Likelihood estimator has the smallest Mean Squared Error of any estimator of p (no bias and smallest possible variance.) Any other choice of α and/or β will increase the bias and lead to an estimator which is not the Maximum Likelihood estimator and therefore has variance than is at least as large as the one of the Maximum Likelihood estimator. The answer is therefore no: setting $\alpha = \beta = 0$ gives us the Maximum Likelihood estimator which is the estimator of p with the smallest Mean Squared Error in this setting; any other choice for the hyper-parameters leads to an estimator with larger Mean Squared Error.**

Prob.II: One of the important features of the power supply unit of a computer server is to deliver power to the server at a steady voltage, regardless of the demand of the server. You are considering placing a large order of power units for a cluster of servers that your company is setting up but, since these are quite expensive, you want to know if they work as advertised.

To do this, you purchased one power unit and carefully measured the output voltage every hour, for one day. Let $X_1, \dots X_{24}$ be independent random variables modelling the voltage output (in volt) for each of the 24 hourly measurements. It can safely be assumed these have a Normal distribution with unknown mean μ, and standard deviation $\sigma = 0.25$ volt.

Suppose that these units are advertised to have an output of 5 volt; you are quite confident that the output is not above 5 volt. You are not sure, however, if it might happen that the voltage might drop below 5 volt – if this were to happen, then the server might shut down, with dramatic consequences.

(To answer the following questions you may need one or more of the following quantiles: $z_{0.01} = -2.33$, $z_{0.0125} = -2.24$, $z_{0.025} = -1.96$, $z_{0.05} = -1.64$.)

(a) You decide to base your purchase decision on the outcome of a statistical test. You want to make sure that you set up the test such that there is a small chance α that the test will tell you not to buy the power units in the case where the units work as advertised. Formulate the appropriate null hypothesis and alternative hypothesis for such a test. **Solution: Based on the description, rejecting means "not buying" and the only situation where we don't buy is when the voltage is too low (meaning lower than the advertised 5 volt.) This means that the correct hypotheses to consider are $H_0 : \mu = 5$ vs $H_1 : \mu < 5$ (the null corresponds to what is advertised and the alternative is what would give us a reason not to buy.)**

(b) Report an appropriate **test statistic** to be used and the **rejection rule** that ensures that the test has the right significance level. **Calibrate the test** to have significance level $\alpha = 0.05$. **Solution: As test statistic we can use $T = \bar{X}$ and, based on the alternative, we should reject if $T \le c^*$ for some appropriate c^*. We need to pick c^* so that under the null hypothesis (i.e., when $\mu = 5$ volt) we reject H_0 with probability α. Since under the null $T = \bar{X} \sim N(5, 0.25^2/n)$, then**

$$\alpha = \mathbb{P}_{\mu=5}(T \le c^*) = \mathbb{P}_{\mu=5}\left(\sqrt{n}\frac{T-5}{0.25} \le \sqrt{n}\frac{c^*-5}{0.25}\right) = \Phi\left(\sqrt{n}\frac{c^*-5}{0.25}\right).$$

This means we want $\sqrt{n}(c^* - 5)/0.25 = \Phi^{-1}(\alpha) = z_\alpha$, which solving for c^* gives $c^* = 5 + 0.25 z_\alpha/\sqrt{n}$. Plugging in $\alpha = 0.05$ so that $z_\alpha = -1.645$, and setting $n = 24$, we get $c^* = 4.9161$. So the test that rejects H_0 when $T \le 4.9161$ has level 0.05.

(c) From the measurements that you took, you computed a voltage sample mean of $\bar{x} = 4.91$ volt. Show that the p-value of the test is $\Phi(-1.7636) = 0.0389$. Should you reject the null hypothesis at significance level $\alpha = 0.05$? **Solution: The p-value is the smallest significance level at which we would reject so it is the smallest α such that $4.91 = t \le c^* = 5 + 0.25 z_\alpha/\sqrt{24}$. Solving this for α we see that we reject if $\sqrt{24}(4.91 - 5)/0.25 \le z_\alpha = \Phi^{-1}(\alpha)$, or if $\alpha \ge \Phi(\sqrt{24}(4.91 - 5)/0.25) = \Phi(-1.7636) = 0.0389$; we conclude that the smallest α that leads to rejection is 0.0389 and this is our p-value. Since the p-value is smaller than 0.05, the conclusion is that we should reject H_0 (and thus not buy the power adapters.)**

(d) Repeat question (c), for a significance level $\alpha = 0.01$. In other words, compute the p-value and make a decision when $\alpha = 0.01$. **Solution: The value of the test statistic and the p-value obviously don't change. The only thing that changes is that now the p-value is greater than 0.01 and so we do not reject the null.**

(e) Suppose that actually $\mu = 4.95$ volt. Is the power of the test in (b) at least 0.5? (Justify your answer by computing the power.) **Solution: The power when $\mu = 4.95$ is the probability of rejecting H_0 when $\mu = 4.95$. We reject H_0 if $T \leq 4.9161$, so the power in this case is $\mathbb{P}_{\mu=4.95}(T \leq 4.9161) = \mathbb{P}_{\mu=4.95}(\sqrt{24}(T-4.95)/0.25 \leq \sqrt{24}(4.9161-4.95)/0.25) = \Phi(-0.6651)$. This is less than 0.5 since Φ is increasing, $-0.6651 < 0$, and $\Phi(0) = 1/2$.**

Prob.III: A common background sound in many places during Summer when you go for a walk in nature is the sound of field crickets. You may have noticed before that crickets tend to chirp (sing) faster when it's warmer. In Table 1 you can see some data (x_i, Y_i), $i = 1, \ldots, 42$: the x_i represents the number of times that a cricket chirped in 60 seconds, and Y_i represents the respective temperature (in degrees Celsius) at the time that you counted the chirps. We plot the data in Figure 1.

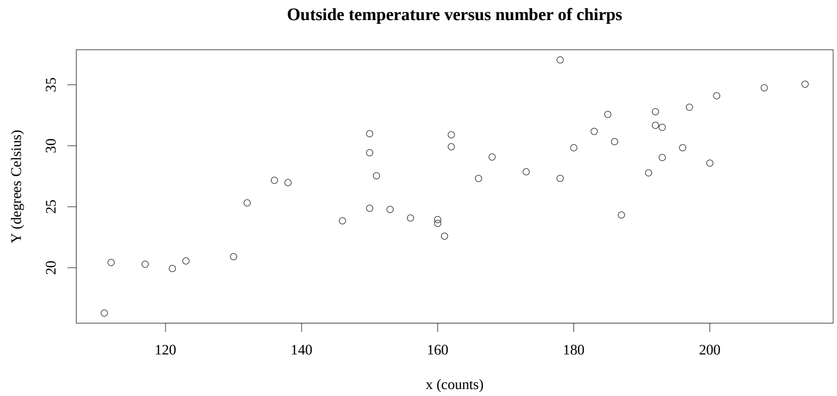

Figure 1: Plot of the cricket dataset.

The data in Figure 1 seems to suggest that the relation between the number of times that a cricket chirps and the outside temperature might be linear. The actual data follows in Table 1.

Variables	Values
$(x_1, \ldots, x_n)$	111 112 117 121 123 130 132 136 138 146 150 150 150 151 153 156 160 160 161 162 162 166 168 173 178 178 180 183 185 186 187 191 192 192 193 193 196 197 200 201 208 214
$(y_1, \ldots, y_n)$	16.29 20.43 20.29 19.94 20.56 20.91 25.32 27.17 26.98 23.85 29.43 24.88 30.99 27.54 24.78 24.08 23.64 23.95 22.59 30.90 29.92 27.32 29.08 27.87 37.04 27.33 29.84 31.18 32.57 30.34 24.33 27.78 32.79 31.68 29.04 31.51 29.84 33.16 28.58 34.10 34.75 35.05

Table 1: The cricket dataset.

From the observations in Table 1 we see that $n\bar{x} = 6942.00$, $n\bar{y} = 1159.60$, $SS_{xx} = 31894.57$, and $SS_{xy} = 4355.36$. There are 42 measurements in total.

(a) Suppose that you would like to use a Simple Linear Regression model to derive a formula that allows you to predict the outside temperature (Y) based on the number of times that a cricket chirps in 60 seconds (x). In a linear regression model you assume that

$$Y_i = \alpha + \beta\, x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\alpha, \beta \in \mathbb{R}$ are unknown, and the ϵ_i are random error terms. In order for Simple Linear Regression to be an adequate model here, what should you assume about: (i) the relation between (x_i, Y_i) and (x_j, Y_j), $i \neq j$; (ii) the expectation of the noise terms ϵ_i; (iii) the variance of the noise terms ϵ_i.

Solution: (i) These should be independent; (ii) the noise should have expectation 0; (iii) the variance should not depend on i.

(b) Consider the data from Table 1 and assume that the assumption from (a) hold. Based on the data, what are your estimates of the intercept and the slope of the line in your model? (If you do not manage to compute the estimates, assume in the following that your prediction formula is $\hat{Y} = 7.82 + 0.11\,x$.)

Solution: We have that $\hat{\beta} = S_{xY}/SS_{xx} = 4355.36/31894.57 = 0.1366$ and $\hat{\alpha} = \bar{Y} - \bar{x} \times SS_{xY}/SS_{xx} = \bar{Y} - \bar{x} \times \hat{\beta} = 1159.60/42 - 6942.00/42 \times 0.1366 = 5.0378$.

(c) It might just be that there is actually no relation between x and Y. In this case you would expect x not to help explain Y or, in other words, you would expect the slope β to be 0. To test this out assume that the noise terms have a Normal distribution with standard deviation $\sigma = 2$. Based on the fact that you know that in this case

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{SS_{xx}}\right),$$

derive an exact, 95% confidence interval for β. Based on this, what do you conclude about the possibility that the slope might be 0? (You may need one or more of the following Normal quantiles: $z_{0.01} = -2.33$, $z_{0.0125} = -2.24$, $z_{0.025} = -1.96$, $z_{0.05} = -1.64$.)

Solution: From the distribution of $\hat{\beta}$ we see that $Z = (\hat{\beta} - \beta)/(\sigma/\sqrt{SS_{xx}}) \sim N(0,1)$ is a pivot. This means that $\mathbb{P}(z_{0.025} \leq Z \leq z_{1-0.025}) = 1 - 0.05 = 0.95$. The event $\{z_{0.025} \leq Z \leq z_{1-0.025}\}$ is the same as $\{z_{0.025} \leq Z \leq -z_{0.025}\}$ which is the same as $\{z_{0.025} \leq (\hat{\beta} - \beta)/(\sigma/\sqrt{SS_{xx}}) \leq -z_{0.025}\}$ which in turn is the same as $\{\hat{\beta} + \sigma z_{0.025}/\sqrt{SS_{xx}} \leq \beta \leq \hat{\beta} - \sigma z_{0.025}/\sqrt{SS_{xx}}\}$. We conclude that $\mathbb{P}(\hat{\beta} + \sigma z_{0.025}/\sqrt{SS_{xx}} \leq \beta \leq \hat{\beta} - \sigma z_{0.025}/\sqrt{SS_{xx}}) = 0.95$, so that $[\hat{\beta} + \sigma z_{0.025}/\sqrt{SS_{xx}}, \hat{\beta} - \sigma z_{0.025}/\sqrt{SS_{xx}}]$ is a confidence interval of β of level 0.95. Plugging in $\sigma = 2$, $z_{0.025} = -1.96$, and the values of $\hat{\beta}$ and SS_{xx}, we get $[0.1146, 0.1585]$ which does not contain 0, so we can be quite confident that $\beta \neq 0$.

(d) Say that you go outside and hear a cricket chirp 181 times in 60 seconds. What would be your prediction of the outside temperature?

Solution: Using our prediction formula $\hat{Y} = \hat{\alpha} + \hat{\beta} \times x = 5.0378 + 0.1366 \times x$ and plugging in $x = 181$ we get the prediction $\hat{Y} = 29.7560$. (Using the prediction formula $\hat{Y} = 7.82 + 0.11 \times x$ you would get $\hat{Y} = 27.73$.)