STAT 512 sp 2020 Exam II

Karl B. Gregory

This is a take-home test due to the COVID-19 suspension of face-to-face instruction. Do not communicate with classmates about the exam until after its due date/time. You may

- *Use your notes and the lecture notes.*
- *Use books.*
- *NOT work together with others.*

Write all answers on blank sheets of paper; then take pictures and merge to a PDF. Upload a single PDF to Blackboard.

pmf/pdf	$\mathcal{X}$	$M_X(t)$	$\mathbb{E}X$	$\text{Var}\, X$
$p_X(x;p) = p^x(1-p)^{1-x}$,	$x = 0, 1$	$pe^t + (1-p)$	p	$p(1-p)$
$p_X(x;n,p) = \binom{n}{x}p^x(1-p)^{n-x}$,	$x = 0, 1, \dots, n$	$[pe^t + (1-p)]^n$	np	$np(1-p)$
$p_X(x;p) = (1-p)^{x-1}p$,	$x = 1, 2, \dots$	$\frac{pe^t}{1-(1-p)e^t}$	p^{-1}	$(1-p)p^{-2}$
$p_X(x;p,r) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$,	$x = r, r+1, \dots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	rp^{-1}	$r(1-p)p^{-2}$
$p_X(x;\lambda) = e^{-\lambda}\lambda^x/x!$	$x = 0, 1, \dots$	$e^{\lambda(e^t-1)}$	λ	λ
$p_X(x;N,M,K) = \binom{M}{x}\binom{N-M}{K-x}/\binom{N}{K}$	$x = 0, 1, \dots, K$	¡complicadísimo!	$\frac{KM}{N}$	$\frac{KM}{N}\frac{(N-K)(N-M)}{N(N-1)}$
$p_X(x;K) = \frac{1}{K}$	$x = 1, \dots, K$	$\frac{1}{K}\sum_{x=1}^K e^{tx}$	$\frac{K+1}{2}$	$\frac{(K+1)(K-1)}{12}$
$p_X(x;x_1,\dots,x_n) = \frac{1}{n}$	$x = x_1, \dots, x_n$	$\frac{1}{n}\sum_{i=1}^n e^{tx_i}$	$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$	$\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$
$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$-\infty < x < \infty$	$e^{\mu t + \sigma^2 t^2/2}$	μ	σ^2
$f_X(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$0 < x < \infty$	$(1-\beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
$f_X(x;\lambda) = \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right)$	$0 < x < \infty$	$(1-\lambda t)^{-1}$	λ	λ^2
$f_X(x;\nu) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}}x^{\nu/2-1}\exp\left(-\frac{x}{2}\right)$	$0 < x < \infty$	$(1-2t)^{-\nu/2}$	ν	2ν
$f_X(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$0 < x < 1$	$1 + \sum_{k=1}^\infty \frac{t^k}{k!}\left(\prod_{r=0}^k \frac{\alpha+r}{\alpha+\beta+r}\right)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}[F_X(x)]^{k-1}[1-F_X(x)]^{n-k}f_X(x)$$

1. Copy down this sentence on your answer sheet and put your signature underneath: *I have not collaborated with any other student on this exam. The work I have presented is my own.*

2. Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Gamma}(2, \beta)$.

   (a) Write down the pdf of the Gamma$(2, \beta)$ distribution.

   Solution: The pdf of the Gamma$(2, \beta)$ distribution is given by
   $$f_X(x) = \frac{1}{\Gamma(2)\beta^2} x^{2-1} e^{-x/\beta} = \frac{1}{\beta^2} x e^{-x/\beta}, \quad x > 0.$$

   (b) Give $\mathbb{E}S_n^2$.

   Solution: We know that S_n is an unbiased estimator of the population variance, which is for the $\Gamma(2, \beta)$ distribution equal to $2\beta^2$. So
   $$\mathbb{E}S_n^2 = 2\beta^2.$$

   (c) Give $\mathbb{E}\bar{X}_n$.

   Solution: We know that $\bar{X}_n$ is an unbiased estimator of the population mean, which is for the $\Gamma(2, \beta)$ distribution equal to 2β. So
   $$\mathbb{E}\bar{X}_n = 2\beta.$$

   (d) Give $\operatorname{Var} \bar{X}_n$.

   Solution: The variance of the sample mean is equal to the population variance divided by n. Therefore
   $$\operatorname{Var} \bar{X}_n = 2\beta^2/n.$$

   (e) Propose an unbiased estimator $\hat{\beta}$ of β which is based on $\bar{X}_n$.

   Solution: The estimator $\hat{\beta} = \bar{X}_n/2$ has expected value
   $$\mathbb{E}\hat{\beta} = \mathbb{E}\bar{X}_n/2 = 2\beta/2 = \beta,$$
   so it is an unbiased estimator of β.

   (f) Find the variance of your estimator $\hat{\beta}$.

Solution: We have

$$\operatorname{Var} \hat{\beta} = \operatorname{Var}[\bar{X}_n/2] = (1/4) \operatorname{Var} \bar{X}_n = (1/4)2\beta^2/n = \beta^2/(2n).$$

(g) Find $\operatorname{MSE} \hat{\beta}$.

Solution: Since $\hat{\beta}$ is an unbiased estimator of β, $\operatorname{MSE} \hat{\beta} = \operatorname{Var} \hat{\beta} = \beta^2/(2n)$.

(h) Argue whether your estimator $\hat{\beta}$ is consistent for β.

Solution: Since $\operatorname{MSE} \hat{\beta} = \beta^2/(2n) \to 0$ as $n \to \infty$, $\hat{\beta}$ is a consistent estimator of β.

3. Let $\hat{\theta}_n$ be an estimator of a parameter θ.

(a) Explain in your own words what it means if $\hat{\theta}_n$ is a consistent estimator for θ.

Solution: If $\hat{\theta}_n$ is a consistent estimator for θ, it means that it tends to take values closer and closer to θ as the sample size n is increased; collecting more data will help the estimator to be closer to its target.

(b) Explain how you would go about checking whether $\hat{\theta}_n$ is a consistent estimator for θ.

Solution: You can compute its mean squared error MSE and show that it goes to 0 as n goes to infinity, or you can use the definition of consistency and directly show

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1.$$

(c) Explain in your own words what it means if $\hat{\theta}_n$ is a biased estimator of θ.

Solution: If $\hat{\theta}_n$ is an unbiased estimator of θ it means that if data were collected many times and $\hat{\theta}$ were computed on these many data sets, the average of all the $\hat{\theta}$ values should be very close to θ. In other words, the estimator is properly "centered" at θ.

(d) Suppose the variance of $\hat{\theta}_n$ does not get smaller as n is increased. What does this mean in terms of whether the estimator is consistent?

Solution: If the variance of $\hat{\theta}_n$ does not get smaller as $n \to \infty$, then the MSE cannot go to zero as $n \to \infty$, so the estimator cannot be consistent.

4. Let $X_1, \ldots, X_n$ be a random sample from some distribution with mean μ and variance $\sigma^2 < \infty$.

(a) Find $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < 2\sigma/\sqrt{n})$. You may use R to compute any probabilities (just write by hand any R code you used).

> **Solution:** We have
>
> $$\begin{aligned}\lim_{n\infty} P(|\bar{X}_n - \mu| < 2\sigma/\sqrt{n}) &= \lim_{n\to\infty} P\left(-2 < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < 2\right) \\ &= P(-2 < Z < 2), \quad Z \sim \text{Normal}(0,1).\end{aligned}$$
>
> This is given by $\Phi(2) - \Phi(-2) =$ `pnorm(2) - pnorm(-2)` $= 0.9544997$.

(b) What theorem did you invoke in order to get your answer to the previous question?

> **Solution:** The central limit theorem allowed us to treat $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ as a Normal$(0,1)$ after taking the limit as $n \to \infty$.

(c) Find $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < 0.001 \cdot \sigma)$.

> **Solution:** Since $\bar{X}_n$ is a consistent estimator for μ, this limit is equal to 1.

(d) Is the probability $P(\mu - 1 < \bar{X}_n < \mu + 1)$ an increasing or a decreasing function of n? Explain your answer.

> **Solution:** Since $\bar{X}_n$ is a consistent estimator of μ, the probability that $\bar{X}_n$ falls within any fixed distance of μ approaches 1 as the sample size n grows. Therefore it is an *increasing* function of n.

(e) Give the form of a large-sample 98% confidence interval for μ.

> **Solution:** A large-sample 98% confidence interval for μ is given by
>
> $$\bar{X}_n \pm z_{0.02/2} \cdot S_n/\sqrt{n} = \bar{X}_n \pm 2.326348 \cdot S_n/\sqrt{n},$$
>
> where $z_{0.02/2} =$ `qnorm(1 - 0.02/2)` $= 2.326348$.

5. Let X_1, X_2, X_3 be independent rvs with the Uniform$(\theta - 1, \theta + 1)$ distribution, with density given by

$$f_X(x) = \frac{1}{2}\mathbf{1}(\theta - 1 < x < \theta + 1).$$

Consider the three estimators of θ given by

$$\begin{aligned}\hat{\theta} &= (X_1 + X_2 + X_3)/3 \\ \tilde{\theta} &= X_{(2)} \\ \breve{\theta} &= (X_{(1)} + X_{(3)})/2,\end{aligned}$$

where $X_{(1)} < X_{(2)} < X_{(3)}$ are the ordered values of X_1, X_2, X_3.

(a) Give $\mathbb{E}\hat{\theta}$.

Solution: We have $\mathbb{E}\hat{\theta} = \mathbb{E}\bar{X}_3 = \theta$, since the sample mean is an unbiased estimator of the population mean; the mean of the Uniform$(\theta - 1, \theta + 1)$ distribution is $(\theta - 1 + \theta + 1)/2 = \theta$.

(b) Using the fact that the variance of the Uniform(a, b) distribution is $(b - a)^2/12$, give $\operatorname{Var}\hat{\theta}$.

Solution: The population variance is given by

$$\frac{((\theta+1) - (\theta - 1))^2}{12} = 4/12 = 1/3,$$

and the variance of the sample mean is the population variance divided by the sample size, so we have

$$\operatorname{Var}\hat{\theta} = \bar{X}_3 = (1/3)/3 = 1/9.$$

(c) Give $\operatorname{MSE}\hat{\theta}$.

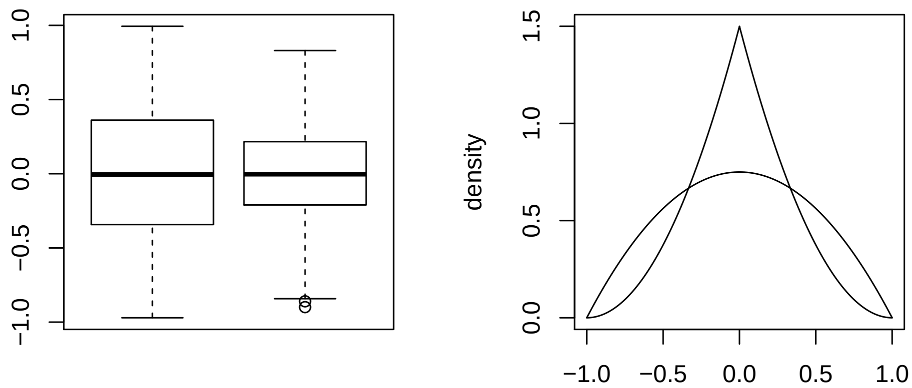
Solution: Since $\hat{\theta}$ is unbiased, $\operatorname{MSE}\hat{\theta} = \operatorname{Var}\hat{\theta} = 1/9$.

(d) Find the pdf of $X_{(2)}$.

Solution: We have

$$\begin{aligned} f_{X_{(2)}}(x) &= \frac{3!}{(2-1)!(3-2)!}\left[\frac{x - (\theta - 1)}{2}\right]^{2-1}\left[1 - \frac{x - (\theta - 1)}{2}\right]^{3-2}\frac{1}{2} \\ &= 3\left[\frac{x - (\theta - 1)}{2}\right]\left[1 - \frac{x - (\theta - 1)}{2}\right]^{3-2} \\ &= \frac{3}{4}\left[1 - (x - \theta)^2\right], \quad \theta - 1 < x < \theta + 1. \end{aligned}$$

(e) It turns out that $\operatorname{Var}\tilde{\theta} = 1/5$ and $\operatorname{Var}\check{\theta} = 1/10$. A simulation was run in which 1,000 random samples of size 3 were drawn from the Uniform$(\theta - 1, \theta + 1)$ distribution under $\theta = 0$. The left panel of the figure shows boxplots of the 1,000 values of $\tilde{\theta}$ and $\check{\theta}$ and the right panel shows plots of the pdfs of these estimators under $\theta = 0$.



i. Comment on whether you think the estimators $\tilde{\theta}$ and $\check{\theta}$ are biased or unbiased.

Solution: Since the boxplots in from the simulation are both centered at $\theta = 0$, it appears that the estimators are unbiased. Moreover, the densities of the estimators $\tilde{\theta}$ and $\check{\theta}$ are both symmetric around $\theta = 0$, so that if we were to compute their expected values, we would find them to be equal to $\theta = 0$.

ii. Of the three estimators $\hat{\theta}$, $\tilde{\theta}$, and $\check{\theta}$, choose the one you think is the best. Explain why you think it is the best.

Solution: If we prefer the estimator with the smallest MSE, we would choose $\check{\theta}$, since all three estimators are unbiased and $\check{\theta}$ has the smallest variance.