5.2.15

The exponential pdf is a measure of lifetimes of devices that do not age (see Question 3.11.11). However, the exponential pdf is a special case of the Weibull distribution, which can measure time to failure fo devices where the probability of failure increases as time does. A Weibull random variable Y has pdf $f_Y(y; \alpha, \beta) = \alpha \beta y^{\beta-1} e^{-\alpha y^\beta}, 0 \leq y, 0 < \alpha, 0 < \beta$. Find the maximum likelihood estimator for α assuming that β is known.

To find the MLE, we can go through the general procedure of taking the likelihood function given the pdf and then try to take the derivative and set it to zero, also seeing if taking the log before doing so would be helpful. Let's begin by taking the likelihood function for the continuous Y:

$$L(\theta) = \prod_{i=1}^{n} f_Y(y_i; \theta) = L(\alpha, \beta) = \prod_{i=1}^{n} f_Y(y_i; \alpha, \beta) = \prod_{i=1}^{n} \alpha \beta y_i^{\beta-1} e^{-\alpha y_i^\beta}$$

$$= (\alpha\beta)^n \prod_{i=1}^{n} y_i^{\beta-1} e^{-\alpha y_i^\beta} = (\alpha\beta)^n y_i^{\beta-1} e^{-\alpha \sum_{i=1}^{n} y_i^\beta}$$

We can then take the natural logarithm of the likelihood to help simplify a bit:

$$\ln L(\alpha, \beta) = \ln[(\alpha\beta)^n \prod_{i=1}^{n} y_i^{\beta-1} e^{-\alpha \sum_{i=1}^{n} y_i^\beta}] = n \ln \alpha + n \ln \beta + (\beta-1) \sum_{i=1}^{n} \ln y_i - \alpha \sum_{i=1}^{n} y_i^\beta$$

We can then take the derivative of this function and set to zero. As β is known, we only need to take the derivative with respect to α. To calculate the estimator, we therefore only need to care about the terms with α, so we can simplify to

$$= n \ln \alpha - \alpha \sum_{i=1}^{n} y_i^\beta$$

And then take the derivative equal to zero:

$$\frac{d}{d\alpha}(n \ln \alpha - \alpha \sum_{i=1}^{n} y_i^\beta) = 0$$

$$n \cdot \frac{1}{\alpha} - \sum_{i=1}^{n} y_i^{\beta} = 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} y_i^{\beta}$$

$$\alpha = \frac{n}{\sum_{i=1}^{n} y_i^{\beta}}$$

So we find that $\boxed{\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} y_i^{\beta}}}$

5.4.4

A sample of size of $n = 16$ is drawn from a normal distribution where $\sigma = 10$ but μ is unknown. If $\mu = 20$, what is the probability that the estimator $\hat{\mu} = \overline{Y}$ will lie between 19.0 and 21.0?

We are estimating using the sample mean. We have $E(\overline{Y}) = \mu = 20$ and $Var(\overline{Y}) = \frac{\sigma^2}{n} = \frac{10^2}{16} = \frac{100}{16}$, the latter of which gives us $\sigma = \frac{10}{4} = \frac{5}{2}$. By definition, we are given a distribution $N(\mu, \sigma^2) = N(20, \frac{5}{2})$.

As we want to find $P(19.0 < \overline{Y} < 21.0)$, we can do this by standardizing according to our distribution:

$$P(19.0 < \overline{Y} < 21.0) = P(\frac{19.0 - 20}{\frac{5}{2}} < Z < \frac{21.0 - 20}{\frac{5}{2}} = P(-\frac{2}{5} < Z < \frac{2}{5}))$$

So we take from the table, looking for values corresponding to 0.4 and -0.4, finding this to be 0.65542 and 0.34458 respectively. We can thus calculate this as 0.65542 - 0.34458 = 0.31084, so the probability is approximately $\boxed{31.08\%}$

5.4.16

Is the maximum likelihood estimator for σ^2 in a normal pdf, where both μ and σ^2 are unknown, asymptotically unbiased?

Let's first derive the MLE for σ^2 using the standard method of taking the likelihood function, the natural logarithm of it, and the derivative with respect to σ^2. Let's begin using the definition of the normal pdf, and take the likelihood function:

$$L(\theta) = L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{X_i-\mu}{\sigma})^2}$$

$$= (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{\sum_{i=1}^{n} -\frac{1}{2}(\frac{X_i-\mu}{\sigma})^2} = (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}$$

Let's then take the natural logarithm of this in the same way:

$$\ln L(\mu, \sigma^2) = \ln[(\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}]$$

$$= -\frac{1}{2} \cdot n \ln 2\pi - \frac{1}{2} \cdot n \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2$$

$$= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2$$

Since we are finding the MLE, we can just keep the unknown terms including the terms we are optimizing for, so in the context of looking for the MLE for σ^2 we can remove the first constant term and are left with $-\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2$. We then take the derivative of this with respect to σ^2 and then set equal to zero:

$$\frac{\partial}{\partial \sigma^2}[-\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2] = 0$$

$$-\frac{n}{2} \cdot \frac{1}{\sigma^2} - [\frac{1}{2} \sum_{i=1}^{n}(X_i - \mu)^2] \cdot -\frac{1}{(\sigma^2)^2} = 0$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

$$\frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n}(X_i - \mu)^2 = \frac{n}{2\sigma^2}$$

$$\frac{1}{\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2 = n$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n}(X_i - \mu)^2$$

So therefore we have $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. We can now look at this MLE in regards to it being asymptotically unbiased. By definition, we know something is asymptotically unbiased if $\lim_{n\to\infty} E(Y'_n) = \theta$, where in this case $Y'_n = \hat{\sigma^2}$ and $\theta = \sigma^2$. In other words, we aim to determine whether $\lim_{n\to\infty} E(\hat{\sigma^2}) = \sigma^2$. We note our estimator is nothing but the definition of the sample variance multiplied by $\frac{n-1}{n}$, which means that we have

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

which taken asymptotically as n approaches infinity is simply equivalent to σ^2. So $\boxed{\text{yes}}$ – this indeed does fulfill the criteria for being asymptotically unbiased.

5.5.2

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the Poisson distribution, $p_X(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, \ldots$. Show that $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}X_i$ is an efficient estimator for λ.

By definition, the efficient estimator is said to be the unbiased estimator with variance equal to the Cramer-Rao lower bound. We will perform each step separately. To show unbiasedness, we need to first show that $E(\hat{\lambda}) = \lambda$:

$$E(\hat{\lambda}) = E(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n}\sum_{i=1}^{n}E(X_i)$$

And since this is a Poisson distribution, we know that $E(X) = \lambda$, so we can directly substitute and continue:

$$= \frac{1}{n}\sum_{i=1}^{n}\lambda = \frac{1}{n}\cdot n \cdot \lambda = \lambda$$

which thus proves unbiasedness. We then need to show that this estimator fulfills the Cramer-Rao lower bound. We can calculate the variance of the estimator as

$$Var(\hat{\lambda}) = Var(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)$$

By definition, we also know the variance of the Poisson distribution to be λ. Substituting accordingly:

$$= \frac{1}{n^2} \sum_{i=1}^{n} \lambda = \frac{1}{n^2} \cdot n \cdot \lambda = \frac{n \cdot \lambda}{n^2} = \frac{\lambda}{n}$$

which is our term for variance. Let's now check this against the Cramer-Roa lower bound using the definition. Using the method in the textbook, we first take the natural log of the probability function:

$$\ln p_X(X_i; \lambda) = \ln(\frac{e^{-\lambda}\lambda^{X_i}}{X_i!})$$

$$\ln p_X(X_i; \lambda) = -\lambda + X_i \ln \lambda - \ln(X_i!)$$

We then take the derivative:

$$\frac{\partial \ln p_X(X_i; \lambda)}{\partial \lambda} = -1 + \frac{X_i}{\lambda} - 0 = \frac{X_i}{\lambda} - 1$$

and then again:

$$\frac{\partial^2 \ln p_X(X_i; \lambda)}{\partial \lambda^2} = -\frac{X_i}{\lambda^2}$$

We then take the expected value of the second derivative, again using the property of the Poisson that $E(X_i) = \lambda$:

$$E(-\frac{X_i}{\lambda^2}) = -\frac{E(X_i)}{\lambda^2} = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}$$

Finally, we take the Cramer-Rao lower bound to be

$$\frac{1}{-n \cdot -\frac{1}{\lambda}} = \frac{1}{\frac{n}{\lambda}} = \frac{\lambda}{n}$$

which is exactly equivalent to the variance that we found. We've satisfied both conditions and shown that $\hat{\lambda}$ is an efficient estimator for λ.

5.6.1

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the geometric distribution, $p_X(k; p) = (1-p)^{k-1}p, k = 1, 2, \cdots$. Show that $\hat{p} = \sum_{i=1}^{n} X_i$ is sufficient for p.

We are given a well-known pdf in the form of the geometric distribution, so we can try to apply the factorization theorem, which was defined as

$$L(\theta) = \sum_{i=1}^{n} p_X(k_i; \theta) = P_{\hat{\theta}}(\hat{\theta} = \theta_e) \cdot b(k_1, k_2, ..., k_n)$$

the left side of which which we can rewrite in terms of p to be

$$= \prod_{i=1}^{n}(1 - p)^{k-1}p = p^n(1 - p)^{-n+\sum_{i=1}^{n} k_i}$$

which only depends on the sum of the k values, not each individual value k_i. We need to show the $P_{\hat{\theta}}(\hat{\theta} = \theta_e) \cdot b(k_1, k_2, ..., k_n)$ part of the theorem above still. We previously studied that the negative binomial can be written as the sum of r independent geometric (p) random variables. We can translate $P_{\hat{\theta}}(\hat{\theta} = \theta_e)$ to be $\binom{k-1}{r-1}p^r(1-p)^{k-r}$. We have $r = n$ and $k = \sum_{i=1}^{n} k_i = Y$ for simplification reasons, and using the pdf that we know through the geometric distribution, we can rewrite this as

$$P\left(X = \sum_{i=1}^{n} k_i\right) \cdot \frac{1}{\binom{\sum_{i=1}^{n} k_i - 1}{n-1}} = P(X = Y) \cdot \frac{1}{\binom{Y-1}{n-1}}$$

where the first part relies on satisfies the $P_{\hat{\theta}}(\hat{\theta} = \theta_e)$ term which relies on p, and the second part which is independent and does not rely on p, as it is just a figure relative to the number of combinations possible, not the actual probability. We have therefore shown sufficiency for $\hat{p}$.

5.7.1

How large a sample must be taken from a normal pdf where $E(Y) = 18$ in order to guarantee that $\hat{\mu}_n = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ has a 90% probability of lying somewhere in the interval $[16, 20]$? Assume that $\sigma = 5.0$.

To determine sample size, we know that in this case with normal distribution and using the formula relevant to means (as we are calculating absolute intervals along the normal distribution) to be $n = (z_{\alpha/2}\sigma/d)^2$. We know that $90\% = 0.9 = 1 - \alpha$, so we have $\alpha = 0.1$. Centering around $E(Y) = 18$, we therefore need to take $z_{0.1/2} = z_{0.05}$,

with the closest value to this being the value 1.64 at 0.94950 which we look up in a table; conveniently, we have been given 1.645 to be this value in the lecture slides. We are give $\sigma = 5.0$ and d is half the interval width, which is $d = \frac{20-16}{2} = 2$. Plugging all this in, we get

$$n = (\frac{1.645 \cdot 5}{2})^2 = 16.91265625$$

as we need whole number samples, the minimum size sample is therefore $\boxed{17}$.