Introduction to mathematical Statistics **Midterm 1 Solution**

1. An expert witness in a paternity suit testifies that the length (in days) of pregnancy (that is, the time from impregnation to the delivery of the child) is approximately normally distributed with parameter $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short pregnancy indicated by the testimony?

Solution: let $X \sim N(\mu = 270,\ \sigma^2 = 100)$ and $Z \sim N(0, 1)$

P (the woman had a very long or very short pregnancy)

$$= P(X > 290) + P(x < 240) = P(Z > \frac{290 - 270}{10}) + P(Z < \frac{240 - 270}{10})$$

$$= P(Z > 2) + P(Z < -3) = .0228 + .0013 = .0241$$

2. Thanksgiving was coming up and Harvey's Turkey Farm was doing a land-office business. Harvey sold 100 gobblers to Nedicks for their famous Turkey-dogs. Nedicks found that 90 of Harvey's turkeys were in reality peacocks.

(a) Estimate the proportion of peacocks at Harvey's Turkey Farm and find a 95% confidence interval for the true proportion of turkeys that Harvey owns.

(b) How large a random sample should we select from Harvey's Farm to guarantee the length of the 95% confidence interval to be no more than 0.06? (Note: please first derive the general formula for sample size calculation based on the length of the CI for inference on one population proportion, large sample situation. Please give the formula for the two cases: (i) we have an estimate of the proportion and (ii) we do not have an estimate of the proportion to be estimated. (iii) Finally, please plug in the numerical values and obtain the sample size for this particular problem.)

Solution:

(a) This is large sample CI for one population proportion.

We have $\alpha = 0.05$, $\hat{\pi} = 0.1$. The 95% CI is

$$\hat{\pi} \pm 1.96\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = 0.1 \pm 1.96\sqrt{\frac{0.1(1-0.1)}{100}} = 0.1 \pm 0.06 \text{ or } (0.04, 0.16)$$

(b) This is sample size calculation for the estimation of one population proportion. The general formula is derived as follows (also refer to your lecture notes for the derivations of the pivotal quantity, and the CI etc.):

(i) The pivotal quantity for the inference on π is

$Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} \sim N(0,1)$. The 100(1-α)% symmetrical CI for π is derived from

$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1-\alpha$, which yields the CI $\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$. Therefore the length of the CI is: $L = 2 * z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$. Solving for the sample size, we have $n = \frac{(z_{\alpha/2})^2 \hat{\pi}(1-\hat{\pi})}{E^2}$, where E = L/2 is referred to as the maximum error.
(Note: E is also directly defined as $P\left(|\hat{\pi} - \pi| \le E\right) = 1-\alpha$)

(ii) When we have no estimate of the proportion, since simple calculus shows that $\hat{\pi}(1-\hat{\pi}) \le 1/4$, a conservative estimate is $n = \frac{(z_{\alpha/2})^2}{4E^2}$
In the given problem, we have E = 0.03 and α = 0.05.
(i). $\hat{\pi} = 0.1$. $n = \frac{(z_{0.025})^2 \hat{\pi}(1-\hat{\pi})}{E^2} \approx 385$; (ii). $n = \frac{(z_{0.025})^2}{4E^2} \approx 1068$

3. Let X_i, i = 1, …, n, denote the outcome of a series of n independent trials, where X_i = 1 with probability p, and Xi = 0 with probability (1- p). Let $X = \sum_{i=1}^{n} X_i$.
(a). Please derive the method of moment estimator of p.
(b). Please derive the maximum likelihood estimator of p.
(c). Please derive the 100(1-α)% large sample confidence interval for p using the pivotal quantity method. (* Please include the derivation of the pivotal quantity, the proof of its distribution, and the derivation of the confidence interval for full credit.)

Solution:

(a). The population mean is p and the sample mean is $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n}$. Therefore the moment estimator of p is $\hat{p} = \frac{X}{n}$.
(b). The likelihood function is:
$L(p; x_1, \cdots, x_n) = \prod_{i=1}^{n}\left[p^{x_i}(1-p)^{1-x_i}\right] = p^{\sum x_i}(1-p)^{n-\sum x_i}$
The log likelihood is:
$\ln L(p; x_1, \cdots, x_n) = (\sum x_i)\ln p + (n - \sum x_i)\ln(1-p)$
Solving the equation:

$\frac{d \ln L(p; x_1, \cdots, x_n)}{dp} = \frac{\sum x_i}{p} - \frac{(n - \sum x_i)}{1-p} = 0$, we have $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x}{n}$

(c). The population distribution is Bernoulli (p), i.e. $X_i \sim$ Bernoulli(p). Therefore the population mean is p and the population variance is p(1-p). When the sample size n is large, by the central limit theorem, we know that the sample mean follows approximately the normal distribution with its mean being the population mean and its

variance being the population variance divided by n as follows:

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n} \dot{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

Same as in 2 (a), we can show that $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \dot{\sim} N(0,1)$ is a pivotal quantity for the inference on p.

We can use this pivotal quantity to construct the large sample confidence interval for p. Alternatively, we can also use the following pivotal quantity $Z^* = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \dot{\sim} N(0,1)$ to construct the large sample confidence interval as follows.

$$1-\alpha = P\left(-Z_{\frac{\alpha}{2}} \le Z^* \le Z_{\frac{\alpha}{2}}\right) \Rightarrow 1-\alpha = P\left(-Z_{\frac{\alpha}{2}} \le \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le Z_{\frac{\alpha}{2}}\right)$$

$$\Rightarrow 1-\alpha = P\left(\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Therefore the 100(1-α)% large sample confidence interval for p is:

$$\left(\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

4. Let X and Y be random variables with joint pdf

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\},$$

$-\infty < x < \infty$, $-\infty < y < \infty$. Then X and Y are said to have the *bivariate normal distribution*. The joint moment generating function for X and Y is

$$M(t_1,t_2) = \exp\left[t_1\mu_X + t_2\mu_Y + \frac{1}{2}\left(t_1^2\sigma_X^2 + 2\rho t_1 t_2\sigma_X\sigma_Y + t_2^2\sigma_Y^2\right)\right].$$

   (a) Find the marginal pdf's of X and Y;
   (b) Prove that X and Y are independent if and only if ρ = 0. (Here ρ is indeed, the correlation coefficient between X and Y.)
   (c) Find the distribution of $(X+Y)$.

Solution:

(a) $M(t_1,0) = \exp(\mu_X t_1 + 1/2\sigma_X^2 t_1^2) \therefore X \sim N(\mu_X, \sigma_X^2)$
$M(0,t_2) = \exp(\mu_Y t_2 + 1/2\sigma_Y^2 t_2^2) \therefore Y \sim N(\mu_Y, \sigma_Y^2)$

(b) If ρ = 0, then $M(t_1, t_2) = \exp[\mu_X t_1 + \mu_Y t_2 + 1/2(\sigma_X^2 t_1^2 + \sigma_Y^2 t_2^2)] = M(t_1,0)\cdot M(0,t_2)$
Therefore X and Y are independent.
If X and Y are independent, then $M(t_1, t_2) = M(t_1,0)\cdot M(0,t_2)$

$= exp\ [\mu_X t_1+\mu_Y t_2+1/2(\sigma_X^2 t_1^2+\sigma_Y^2 t_2^2)]$ *Therefore* $\rho = 0$.
Note: It is interesting to note that for two random variables with bivariate normal distribution, they are independent if and only if they are uncorrelated.

(c) $M_{X+Y}(t) = e^{t(X+Y)}=e^{tX+tY}$
Recall that $M(t_1, t_2) = e^{t_1X+t_2Y}$, *therefore we can obtain* $M_{X+Y}(t)$ *by setting* $t_1=t_2=t$ *in* $M(t_1, t_2)$. *That is,*
$M_{X+Y}(t) = M(t,t)=exp[\mu_X t+\mu_Y t+1/2(\sigma_X^2 t^2+2\rho\sigma_X\sigma_Y t^2+\sigma_Y^2 t^2)]$
$= exp\ [(\mu_X+\mu_Y)t+1/2(\sigma_X^2+2\rho\sigma_X\sigma_Y+\sigma_Y^2)\cdot t^2]$
Hence $X+Y \sim N(\mu=\mu_X+\mu_Y,\ \sigma^{2=}\sigma_X^2+2\rho\sigma_X\sigma_Y+\sigma_Y^2)$