Steve Sawin
Math Stats 3352

Problem Set II (Solutions)

Due Feb. 26

All write-ups should be individually done. Please tell me anyone you worked with.

1. [3pt] Use the moment generating function technique to show that the sum $Y = X_1 + X_2$ of two independent binomial variables with X_1 having parameters n_1 and θ and X_2 having parameters n_2 and θ is again a binomial variable with parameters $n_1 + n_2$ and θ. The moment generating functions are X_1 is

$$M_{X_1}(t) = \left[1 + \theta(e^t - 1)\right]^{n_1}$$
$$M_{X_2}(t) = \left[1 + \theta(e^t - 1)\right]^{n_2}$$

so Theorem 7.3 says that

$$M_Y(t) = \left[1 + \theta(e^t - 1)\right]^{n_1} \times \left[1 + \theta(e^t - 1)\right]^{n_2} = \left[1 + \theta(e^t - 1)\right]^{n_1+n_2}$$

which you recognize as the moment generating function of a binomial distribution with parameters $n_1 + n_2$ and θ.

2. [3pt] A random sample of size $n = 16$ is taken from an infinite population with a random variable having a mean of 75 and a standard deviation of 16. Using the Central Limit Theorem, find the probability that the sample mean will fall between 67 and 83. The sample mean has mean $\mu = 75$ and a standard deviation $\sigma/\sqrt{n} = 16/\sqrt{16} = 4$. That means the range asked about in the problem is 2 standard deviations above and below the mean, so the probability is around 95%. But if you type $\text{normdist}(83, 75, 4, 1) - \text{normdist}(67, 75, 14, 1)$ into Excel you get 0.9545.

3. [4pt] Suppose X and Y are independent *discrete* random variables. Show that $F(X)$ and Y are independent random variables as well, where F is any function. Let $Z = F(x)$. Then for any number z call the values of X that get mapped by F to z $x_1, x_2, \ldots, x_r$, so the statement $Z = z$ is the statement $(X = x_1$ or $X = x_2$ or $\cdots$ or $X = x_r$. So by the sum rule

$$P(Z = z) = P(X = x_1) + P(X = x_2) + \cdots + P(X = x_r).$$

By the same token (and the distributive property of AND and OR)

$$\begin{aligned} P(Z = z \text{ and } Y = y) &= P\big[\,(X = x_1 \text{ and } Y = y) \text{ OR } (X = x_2 \text{ and } Y = y) \text{ OR } \\ &\quad \cdots \text{ OR } (X = x_r \text{ and } Y = y)\,\big] \\ &= P\left(X = x_1 \text{ and } Y = y\right) + P\left(X = x_2 \text{ and } Y = y\right) + \\ &\quad \cdots + P\left(X = x_r \text{ and } Y = y\right) \end{aligned}$$

by the distributive property of AND and OR and the sum rule again. But the independence if X and Y means that $P(X = x \text{ and } Y = y) = P(X = x)P(Y = y)$, so

$$
\begin{aligned}
P(Z = z \text{ and } Y = y) &= P(X = x_1)P(Y = y) + P(X = x_2)P(Y = y) + \cdots + P(X = x_r)P(Y = y) \\
&= [P(X = x_1) + P(X = x_2) + \cdots + P(X = x_r)]\, P(Y = y) \\
&= P(Z = z)P(Y = y)
\end{aligned}
$$

and BOOM, they are independent!

4. [3pt] Show that the formula for the sample variance S^2 can be written as

$$S^2 = \frac{(\sum_{i=1}^n X_i^2) - n\overline{X}^2}{n-1}.$$

$$
\begin{aligned}
S^2 &= \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^n \left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)}{n-1} \\
&= \frac{\sum_{i=1}^n X_i^2 - 2\overline{X}\sum_{i=1}^n X_i + n\overline{X}^2}{n-1} \\
&= \frac{\sum_{i=1}^n X_i^2 - 2\overline{X}n\overline{X} + n\overline{X}^2}{n-1} \\
&= \frac{\sum_{i=1}^n X_i^2 - n\overline{X}^2}{n-1}
\end{aligned}
$$

5. [3pt] A sample of size $n = 100$ is taken from an infinite population with an exponential random variable of parameter $\theta = 4$. Using the Central Limit Theorem, what is the probability that the sample mean will exceed 4.5? As we all remember off the top of our heads, an exponential distribution with parameter θ has mean and standard deviation equal to θ. So in this case the mean and s.d. are 4. The sample mean and standard error are therefore 4 and $4/\sqrt{100} = .4$ respectively. By the Central Limit Theorem we can treat this as a normal distribution of the same mean and s.d., so the probability the sample mean will exceed 4.5 is $1 - \text{normdist}(4.5, 4, .4, 1) = 10.56\%$.

6. [4pt] If $X_1, X_2, \ldots, X_n$ are independent random variables each with mean μ, and $\overline{X}$ is their average, show that

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2.$$

Think of the left hand side as measuring how far the values actually are from the mean, and the right hand side writing that error as a sum of two errors: How far they are from their sample mean $\overline{X}$, and how far $\overline{X}$ is from the true mean.

$$\begin{aligned}
\sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_{i=1}^{n} X_i^2 - 2\mu\sum_{i=1}^{n} X_i + n\mu^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\mu\overline{X} + n\mu^2 - n\overline{X}^2 + n\overline{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\mu\overline{X} + n\mu^2 - n\overline{X}^2 + n\overline{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\mu\overline{X} + n\mu^2 - n\overline{X}^2 + n\overline{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + n\overline{X}^2 + \left(\overline{X}-\mu\right)^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}\overline{X}^2 + \left(\overline{X}-\mu\right)^2 \\
&= \sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 + \left(\overline{X}-\mu\right)^2
\end{aligned}$$

Out of 20 points