

14.30 Statistics - Fall 2003

Exam #2 Solutions

Prepared by Eric Moos

1. True, False, or Uncertain:

- (a) False - The event most likely to occur is called the mode. The mean is what we expect to happen "in expectation." In fact, the mean need not be a possible outcome. For example, consider the Bernoulli distribution (e.g., flipping a coin) with probability of success $\frac{1}{2}$. There are two possible outcomes, $X = 0$ and $X = 1$; but the expected value of X is $\frac{1}{2}$.
- (b) True - Given the transformation $Z = z(X)$, we can compute the expectation of the random variable Z in either of two ways, using $z(X)$ and the distribution of X or using Z and the distribution of Z .

$$\int_{-\infty}^{\infty} z(x) f_X(x) dx = E[z(X)] = E[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz$$

- (c) False - Although the correlation and the covariance are related, they are not the same thing. The correlation is a rescaling of the covariance which always lies between -1 and 1 . The exact relationship between the correlation and the covariance is the following:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

- (d) Uncertain - Although this seems like a Poisson process, in which case the answer would have been true, this may not necessarily be a Poisson process. For it to be a Poisson process, the example must satisfy the following assumptions:
- The number of arrivals in every fixed interval of time of length λt has a Poisson process in which the mean is λt ; and
 - The number of arrivals in every two disjoint time intervals are independent.
- (e) Uncertain - If X has the standard normal distribution, then, by definition, it has mean 0 and variance 1. The statement is only true if $\sigma^2 = 1$.
- (f) False - The Chi-Squared distribution has $f(x) = 0$ for $x < 0$ **for all** k . There are no degrees of freedom for which the pdf is non-zero on the non-positive part of the real line.

2. Expectations of random variables:

(a)

$$\begin{aligned} E[aX + b] &\equiv \int_{-\infty}^{\infty} (ax + b) f_X(x) dx = \int_{-\infty}^{\infty} ax f_X(x) dx + \int_{-\infty}^{\infty} b f_X(x) dx \\ &= a \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} f_X(x) dx = aE[X] + b \end{aligned}$$

- (b) Because X and Y are independent, the joint distribution is the product of the marginal distributions.

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Therefore,

$$\begin{aligned}
 E[g(X)h(X)] &\equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y) dx dy \\
 &= \int_{-\infty}^{\infty} h(y)f_Y(y) \left[\int_{-\infty}^{\infty} g(x)f_X(x) dx \right] dy \\
 &= \int_{-\infty}^{\infty} g(x)f_X(x) dx \int_{-\infty}^{\infty} h(y)f_Y(y) dy \\
 &= E[g(X)]E[h(Y)]
 \end{aligned}$$

- (c) Although this fact is easiest to show using the Law of Iterated Expectations, the question specifically asks you to use the formal definition of expectation. In order to prove this, we shall merely use Bayes Rule (relating joint distributions to marginal and conditional distributions) and the definition of the expectation.

$$\begin{aligned}
 E[Z Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} zy f_{ZY}(z,y) dy dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} zy f_{Y|Z}(y|z) f_Z(z) dy dz \\
 &= \int_{-\infty}^{\infty} z f_Z(z) \left[\int_{-\infty}^{\infty} y f_{Y|Z}(y|z) dy \right] dz = \int_{-\infty}^{\infty} z f_Z(z) E[Y|z] dz \\
 &= E[ZE[Y|Z]]
 \end{aligned}$$

- (d) Using the definition of covariance, the Law of Iterated Expectations, and the result from part c,

$$\begin{aligned}
 Cov[Z, E(Y|Z)] &= E[ZE(Y|Z)] - E[Z]E[E(Y|Z)] \\
 &= E[Z Y] - E[Z]E[Y] \\
 &= Cov[Z, Y]
 \end{aligned}$$

3. Moments:

- (a) Using the definition of the expectation, the mean is

$$\begin{aligned}
 E(X) &= \int_0^2 x \frac{3}{8} x^2 dx = \frac{3}{8} \int_0^2 x^3 dx \\
 &= \frac{3}{8} \left(\frac{1}{4} \right) x^4 \Big|_0^2 = \frac{3}{32} (16 - 0) \\
 &= \frac{3}{2}
 \end{aligned}$$

In order to calculate the standard deviation, it is easiest to calculate the variance first. To do

this, we need the expected value of X^2 .

$$\begin{aligned}
 E(X^2) &= \int_0^2 x^2 \frac{3}{8} x^2 dx = \frac{3}{8} \int_0^2 x^4 dx \\
 &= \frac{3}{8} \left(\frac{1}{5} \right) x^5 \Big|_0^2 = \frac{3}{40} (32 - 0) \\
 &= \frac{12}{5} \\
 \text{Var}(X) &= E(X^2) - E(X)^2 \\
 &= \frac{12}{5} - \frac{9}{4} = \frac{48 - 45}{20} = \frac{3}{20}
 \end{aligned}$$

The standard deviation is the square root of the variance.

$$SD(X) = \sqrt{\text{Var}(X)} = \sqrt{\frac{3}{20}}$$

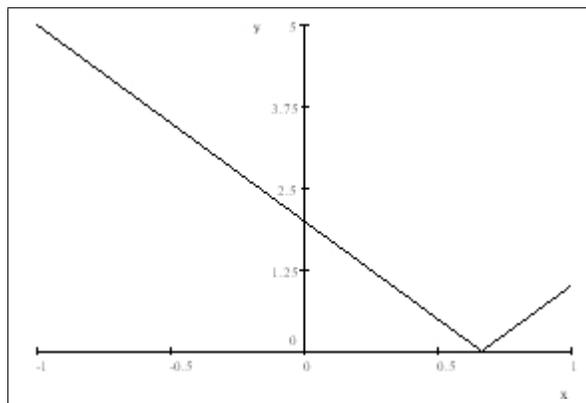
We can expand the variance of a linear function of X in terms of the variance of X .

$$\begin{aligned}
 \text{Var}(3X + 2) &= 9\text{Var}(X) \\
 &= 9 * \frac{3}{20} = \frac{27}{20}
 \end{aligned}$$

- (b) The expectation of Y can be found by multiplying the marginal distribution of X and the conditional distribution of Y to get the joint distribution of X and Y , integrating out X to get the marginal distribution of Y , and taking the expectation of Y . But there is no need to go through all of that since the Law of Iterated Expectations simplifies things considerably. Using what we know about the uniform distribution, (i.e., $Z \sim U[a, b] \Rightarrow E[Z] = \frac{b+a}{2}$).

$$\begin{aligned}
 E(Y) &= E[E(Y|X)] = E\left[\frac{X-1}{2}\right] \\
 &= \frac{1}{2}E[X] - \frac{1}{2} = \frac{1}{2}\left(\frac{3}{2}\right) - \frac{1}{2} \\
 &= \frac{3}{4} - \frac{1}{2} = \frac{1}{4}
 \end{aligned}$$

4. Transformation of uniform random variables: Begin by graphing the transformation in order to get a sense of its monotonicity.



- (a) $X \sim U[-0.5, 0.5]$ Notice that on the interval $[-0.5, 0.5]$, $Y = 2 - 3X$, the transformation is strictly

decreasing, and $Y \in [\frac{1}{2}, \frac{7}{2}]$. Therefore, using the 2-Step method,

$$\begin{aligned} F_Y(y) &\equiv P(Y \leq y) = P(2 - 3X \leq y) \\ &= P\left(X \geq \frac{1}{3}(2 - y)\right) \\ &= 1 - P\left(X \leq \frac{1}{3}(2 - y)\right) \\ &= 1 - \frac{1}{3}(2 - y) - \frac{1}{2} = \frac{y}{3} - \frac{1}{6} \\ f_Y(y) &= \frac{1}{3} \end{aligned}$$

The pdf is obtained by differentiating the cdf. A linear transformation (as this actually turns out to be) of a uniform random variable is also uniform.

- (b) $X \sim U[0, 1]$ Notice that on the interval $[0, 1]$, the transformation is not monotonic. Therefore, in order to use the 1-Step method, we must partition X so that the transformation is monotonic on each interval. On the interval $[0, \frac{2}{3})$ $Y = 2 - 3X$, the transformation is strictly decreasing, and $Y \in (0, 2]$.

$$\begin{aligned} y &= 2 - 3x \\ x &= \frac{1}{3}(2 - y) \\ \left|\frac{dx}{dy}\right| &= \left|\frac{-1}{3}\right| = \frac{1}{3} \end{aligned}$$

Therefore, for this partition,

$$f_Y(y) = f_X\left(\frac{1}{3}(2 - y)\right) \frac{1}{3} = \frac{1}{3}(1) = \frac{1}{3}$$

On the interval $(\frac{2}{3}, 1]$ $Y = 3X - 2$, the transformation is strictly increasing, and $Y \in (0, 1]$.

$$\begin{aligned} y &= 3x - 2 \\ x &= \frac{1}{3}(y + 2) \\ \left|\frac{dx}{dy}\right| &= \left|\frac{1}{3}\right| = \frac{1}{3} \end{aligned}$$

Therefore, for this partition,

$$f_Y(y) = f_X\left(\frac{1}{3}(y + 2)\right) \frac{1}{3} = \frac{1}{3}(1) = \frac{1}{3}$$

Our final answer is the aggregate of these two interval results.

$$f_Y(y) = \begin{cases} \frac{2}{3} & y \in (0, 1] \\ \frac{1}{3} & y \in (1, 2] \\ 0 & \text{otherwise} \end{cases}$$

The pdf at $y = 0$ is unimportant because $P(Y = 0) = 0$.

- (c) There is actually a lot going on in this comment/question. First, a Jacobian is not always necessary. A transformation of a discrete random variable does not require a Jacobian; Jacobians are specific to continuous transformations. Also, if you are transforming continuous random variables using the 2-Step method, you do not need to explicitly calculate a Jacobian, although the Jacobian will end up being taken into consideration implicitly. Last, although you may think that the 1-Step method applied to the transformation of one random variable into another does not require a Jacobian, it does in fact use a Jacobian. Suppose you have the following transformation: $Y = r(X)$. The term $\left|\frac{dr^{-1}(y)}{dy}\right|$ in the 1-Step method is a Jacobian, it just happens to be the Jacobian determinant of a 1×1 (i.e., scalar) matrix.

5. Paternity testing (AKA Life in the NBA): Let X denote the random length in days of pregnancy, $X \sim N(260, \frac{144}{a})$. Also, let Z denote a standard normal random variable ($Z \sim N(0, 1)$).

(a) $a = 4 \Rightarrow X \sim N(260, 36)$

$$\begin{aligned} P(X \leq 270) &= P\left(\frac{X - 260}{6} \leq \frac{270 - 260}{6}\right) \\ &= P\left(Z \leq \frac{5}{3}\right) = P(Z \leq 1.67) \\ &\approx 0.9525 \end{aligned}$$

(b) $a = 4 \Rightarrow X \sim N(260, 36)$

$$\begin{aligned} P(254 < X < 270) &= P\left(\frac{254 - 260}{6} < \frac{X - 260}{6} < \frac{270 - 260}{6}\right) \\ &= P\left(-1 < Z < \frac{5}{3}\right) \\ &= P\left(Z < \frac{5}{3}\right) - P(Z \leq -1) \\ &= P\left(Z < \frac{5}{3}\right) - P(Z > 1) \\ &= P\left(Z < \frac{5}{3}\right) + P(Z \leq 1) - 1 \\ &\approx 0.9525 + 0.8413 - 1 \\ &= 0.7938 \end{aligned}$$

(c) $a = 4 \Rightarrow X \sim N(260, 36)$

$$W = \frac{X}{6} - 43.33 = \frac{X - 260}{6}$$

Hence, $W \sim N(0, 1)$, a standard normal random variable.

$$\begin{aligned} 90\% &= P(-\bar{w} \leq W \leq \bar{w}) \\ &= P(W \leq \bar{w}) - P(W \leq -\bar{w}) \\ &= P(W \leq \bar{w}) - P(W \geq \bar{w}) \\ &= P(W \leq \bar{w}) + P(W \leq \bar{w}) - 1 \\ 1.90 &= 2P(W \leq \bar{w}) \\ P(W \leq \bar{w}) &= 0.95 \\ \bar{w} &\approx 1.65 \end{aligned}$$

(Note that 1.645 is a better estimate, but we shall only use two decimal places.)

(d) The probability that impregnation occurred during the time the "potential" father was away is

$$P(245 \leq X \leq 275)$$

According to the judge's decision rule, the "potential" father will not be declared the father if

$$P(245 \leq X \leq 275) \geq 0.90$$

Hence, the "potential" father will be declared the father if

$$\begin{aligned} 0.90 &\geq P(245 \leq X \leq 275) \\ &= P\left(\sqrt{a}\frac{245 - 260}{12} \leq \sqrt{a}\frac{X - 260}{12} \leq \sqrt{a}\frac{275 - 260}{12}\right) \\ &= P\left(-\sqrt{a}\frac{5}{4} \leq Z \leq \sqrt{a}\frac{5}{4}\right) \end{aligned}$$

This statement looks almost exactly like what we considered in part c.

$$\begin{aligned}\sqrt{a}\frac{5}{4} &\leq 1.65 \\ \sqrt{a} &\leq 1.32 \\ a &\leq 1.74\end{aligned}$$

For relatively small values of a , the variance is large (the larger a is, the smaller the variance). A large variance means that we (and, consequently, the judge) are less sure that the baby was conceived while "potential" father was away. If a is large, then the variance is small and the distribution on X is less disperse; there is much more likelihood that the "potential" father was away when the baby was conceived.