MATH/STAT 394: Probability I
Wellner, 3/14/2000

Final Exam, Solutions

1. (32 points). Suppose that $R \sim$ Poisson(μ), $S \sim$ Poisson(ν), and $T \sim$ Poisson(τ) are all independent. Suppose that $X \equiv R+T$, $Y \equiv S+T$, $Z \equiv R-T$.
(a) Compute $E(X)$, $Var(X)$, $E(Y)$, $Var(Y)$.
(b) Compute $Cov(X,Y)$ and $\rho_{X,Y}$, the correlation between X and Y.
(c) Compute $Cov(X,Z)$ assuming that $\mu = \tau$. Are X and Z independent? (Justify your answer.)
(d) What are the marginal distributions of X and Y?

Solution: (a) $E(X) = E(R+T) = E(R) + E(T) = \mu + \tau$;
$Var(X) = Var(R+T) = Var(R) + Var(T) = \mu + \tau$;
$E(Y) = E(S+T) = E(S) + E(T) = \nu + \tau$;
$Var(Y) = Var(S+T) = Var(S) + Var(T) = \nu + \tau$.
(b) By the properties of Covariance, and since the covariance of independent random variables is zero,

$$\begin{aligned} Cov(X,Y) &= Cov(R+T, S+T) = Cov(R,S) + Cov(R,T) + Cov(T,S) + Cov(T,T) \\ &= 0 + 0 + 0 + Var(T) = Var(T) = \tau\,. \end{aligned}$$

Hence

$$\begin{aligned} \rho_{X,Y} &= \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \\ &= \frac{\tau}{\sqrt{(\mu+\tau)(\nu+\tau)}} \\ &= \frac{1}{\sqrt{(1+(\mu/\tau))(1+(\nu/\tau))}}\,. \end{aligned}$$

This equals 1/2 if $\mu = \nu = \tau$, converges to 0 if $\mu = \nu \to \infty$ with τ fixed (or if $\tau \to 0$ with μ and ν fixed), and converges to 1 if $\mu = \nu \to 0$ for fixed τ (or if $\tau \to \infty$ for fixe μ and ν).
(c)

$$\begin{aligned} Cov(X,Z) &= Cov(R+T, R-T) = Cov(R,R) - Cov(R,T) + Cov(T,R) - Cov(T,T) \\ &= Var(R) - Var(T) = \mu - \tau = 0 \qquad \text{if} \ \ \mu = \tau\,. \end{aligned}$$

But X and Z are not independent! Indeed $X + Z = 2R \ge 0$ so that $Z \ge -X$, and $X - Z = 2T \ge 0$, so that $Z \le X$. Thus $-X \le Z \le X$ with probability 1, and hence for $z > x > 0$, $0 = f_{X,Z}(x,z) \ne f_X(x) f_Z(z) > 0$.
(d) Since R, S, and T are independent Poisson random variables, $X = R + T \sim$ Poisson($\mu + \tau$) and $Y = S + T \sim$ Poisson($\nu + \tau$).

Extra: the joint mass function of (X, Y). Note that the joint pass function of (X, Y) is

$$\begin{aligned}
p(x,y) &= P(X=x, Y=y) = P(R+T=x, S+T=y) \\
&= \sum_{t=0}^{\min\{x,y\}} P(R=x-t, S=y-t, T=t) \\
&= \sum_{t=0}^{\min\{x,y\}} P(R=x-t)P(S=y-t)P(T=t) \\
&= \sum_{t=0}^{\min\{x,y\}} e^{-\mu}\frac{\mu^{x-t}}{(x-t)!}e^{-\nu}\frac{\nu^{y-t}}{(y-t)!}e^{-\tau}\frac{\tau^t}{t!} \\
&= \exp(-\mu-\nu-\tau)\sum_{t=0}^{\min\{x,y\}} \frac{\mu^{x-t}}{(x-t)!}\frac{\nu^{y-t}}{(y-t)!}\frac{\tau^t}{t!}\,.
\end{aligned}$$

This is a "bivariate Poisson" distribution.

2. (20 points) Use the normal approximation to the binomial distribution to find the approximate probability that in 99 tosses of a fair coin there are fewer than 40 heads. Use a continuity correction.

   Solution: Since $T_{99} \sim \text{Binomial}(99, 1/2)$, by the CLT it follows that

$$\begin{aligned}
P(T_{99} < 40) &= P(T_{99} \le 39.5) = P\left(\frac{T_{99} - 99/2}{\sqrt{99(1/2)(1/2)}} \le \frac{39.5 - 99/2}{\sqrt{99(1/2)(1/2)}}\right) \\
&\doteq P(Z \le -2.010) = 1 - .9778 = .0222\,.
\end{aligned}$$

3. (20 points) The event whose probability is found in the preceding problem, "fewer than 40 heads in 99 tosses" can be restated as "100 or more trials are needed to produce the 40th head"; i.e. in the notation of our section on the Bernoulli process, $[T_{99} < 40] = [W_{40} > 99]$. You can approximate the probability of the event on the right side by using the Central Limit Theorem. Do this and compare with the result of the previous problem. Which of the two approximations is likely to be more accurate? (Justify/explain your answer.)

   Solution: Since $W_{40} \sim \text{Negative Binomial}(40, 1/2)$, it follows from the CLT that

$$\begin{aligned}
P(W_{40} > 99) &= P(W_{40} \ge 99.5) \\
&= P\left(\frac{W_{40} - 80}{\sqrt{80}} \ge \frac{99.5 - 80}{\sqrt{80}}\right) \\
&\doteq P(Z \ge 2.180) = 1 - .9854 = 0.0146\,.
\end{aligned}$$

   The approximation based on the binomial distribution will be more accurate because of the symmetry of the binomial distribution. In fact the exact probability is $P(T_{99} < 40) = P(W_{40} > 99) = .02194$.

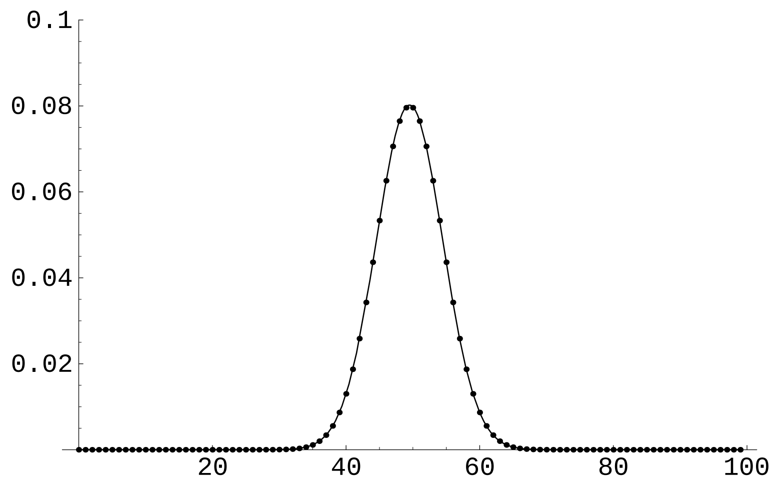

Figure 1: The Binomial$(99, 1/2)$ mass function and normal approximation.

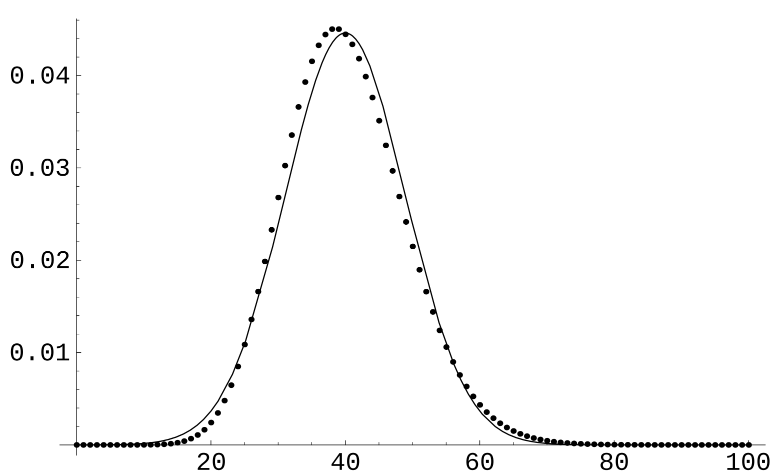

Figure 2: The Negative Binomial$(40, 1/2)$ mass function and normal approximation.

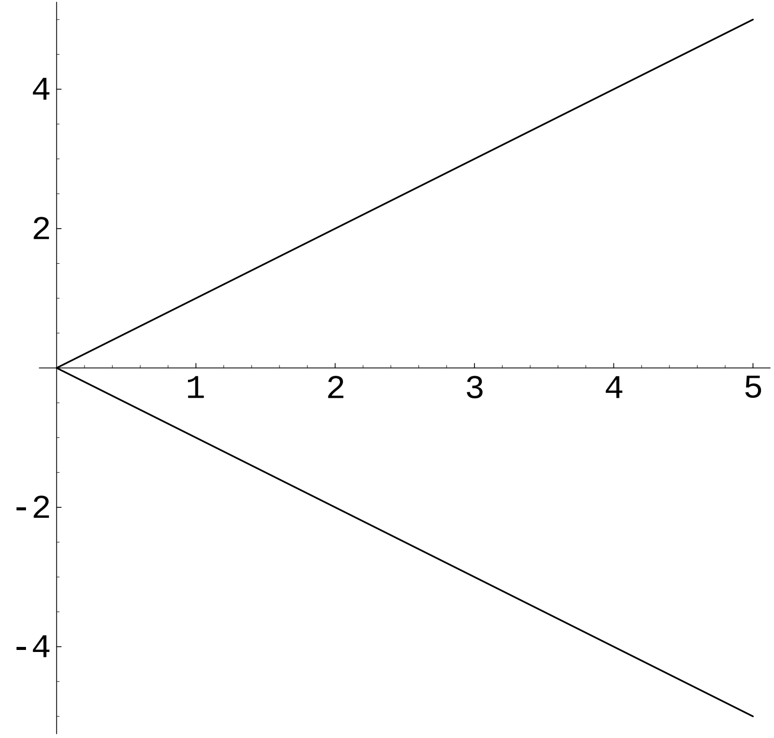

Figure 3: Positivity Region for $f_{X,Y}$ in Problem 4.

4. (32 points) Suppose that the marginal distribution of X is Gamma$(2,\nu)$, and the conditional distribution of Y given $X = x$ is Uniform$(-x, x)$.
(a) Determine the joint density/mass function of X and Y. Show where the joint density is positive in a diagram.
(b) Determine the marginal density of Y. [Hint: consider positive and negative values of y, the argument of the marginal density, separately.]
(c) Are X and Y independent? Justify your answer.
(d) Determine the conditional density of X given $Y = y$.

Solution: (a) Since $f_{Y|X}(y|x) = (1/2x)1_{(-x,x)}(y)$ and $f_X(x) = \nu^2 x \exp(-\nu x)1_{(0,\infty)}(x)$, it follows that the joint density of (X, Y) is given by

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = \frac{\nu^2}{2}e^{-\nu x}1_{(-x,x)}(y)1_{(0,\infty)}(x)\,.$$

This is positive on the region shown in Figure 3.

(b) The marginal density of X is obtained by integrating the joint density over x:

$$f_Y(y) = \int_0^\infty f_{X,Y}(x,y)dx = \int_0^\infty \frac{\nu^2}{2}e^{-\nu x}1_{(-x,x)}(y)1_{(0,\infty)}(x)$$

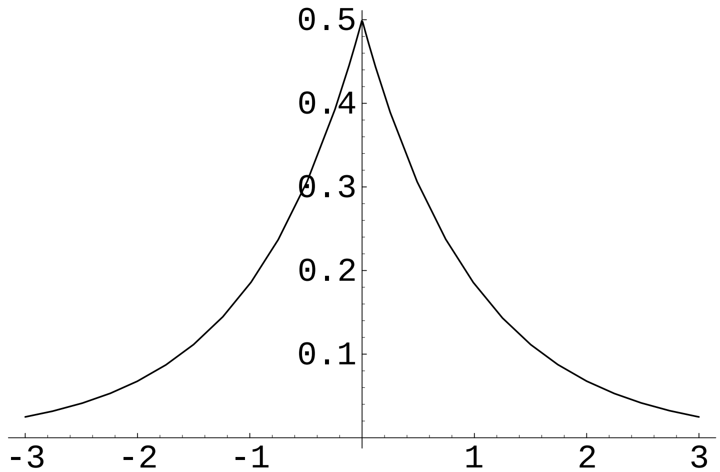

Figure 4: The Laplace(ν) density f_Y with $\nu = 1$.

$$
\begin{aligned}
&= \begin{cases} \int_y^\infty \frac{\nu^2}{2} e^{-\nu x} dx & \text{if } y > 0 \\ \int_{-y}^\infty \frac{\nu^2}{2} e^{-\nu x} dx & \text{if } y \le 0 \end{cases} \\
&= \int_{|y|}^\infty \frac{\nu^2}{2} e^{-\nu x} dx \\
&= \frac{\nu}{2} e^{-\nu|y|} 1_{(-\infty,\infty)}(y) .
\end{aligned}
$$

This is the Laplace(ν) or Double Exponential(ν) density; see Figure 4.

(c) X and Y are not independent since $f_{X,Y}(x,y) = 0 \neq f_X(x) f_Y(y)$ for $y > |x|$.
(d) The conditional density of X given $Y = y$ is given by

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{(\nu^2/2)e^{-\nu x} 1_{(-x,x)}(y) 1_{(0,\infty)}(x)}{(\nu/2)e^{-\nu|y|}} \\
&= \nu \exp(-\nu(x - |y|)) 1_{(|y|,\infty)}(x) .
\end{aligned}
$$

This is a shifted Exponential(ν) density; i.e. $(X - |Y| | Y = y) \sim$ Exponential(ν).

5. (32 points) Customers arrive at the check-out area of an Eagle Hardware store at a rate of 4 per minute on Saturdays in the spring. Break the upcoming hours into 12 disjoint five minute intervals. “Customer overload” is said to occur if such a five minute interval sees the arrival of at least 26 customers.
A. What is the probability of:
(a) Customer overload in the first five minute interval?
(b) The first hour (i.e. sometime in the first 12 five minutes intervals)?
(c) The 2nd customer overload occurring in the 8th interval?

B. Evaluate the mean and variance of the number of intervals until the 6th customer overload.

Solution: A. (a) Now $\lambda = \nu t = 4 \cdot 5 = 20$, so $\mathbb{N}(5) \sim$ Poisson(20). Thus it follows that the probability of an "overload" occurring during the first 5 minutes is

$$\begin{aligned} P(\mathbb{N}(5) \geq 26) &= P(\mathbb{N}(5) \geq 25.5) \\ &= P\left(\frac{\mathbb{N}(5) - 20}{\sqrt{20}} \geq \frac{25.5 - 20}{\sqrt{20}}\right) \\ &\doteq P(Z \geq 1.2298) = 1 - .8907 = .1093 \equiv p\,. \end{aligned}$$

In fact, the exact probability is $P(\mathbb{N}(5) \geq 26) = .1122\ldots$.
(b) Now if we introduce a Bernoulli(p) random variable X_j associated with the jth minute interval, then the total number of overloads in the first hour is $T_{12} = X_1 + \cdots + X_{12} \sim$ Binomial$(12, p)$. The probability of at least one overload during the first hour is (using the approximation from (a)),

$$P(T_{12} \geq 1) = 1 - P(T_{12} = 0) = 1 - (1-p)^{12} = .7507\,.$$

(Thus the exact probability is $1 - (1 - .1122)^{12} = .760\ldots$.)
(c) Now W_2, the number of intervals until the second overload, has a Negative Binomial$(2, p)$ distribution, and hence (using the approximation from (a)),

$$P(W_2 = 8) = \binom{7}{1} q^6 p^2 = 7 \cdot (.8907)^6 \cdot (.1093)^2 = 0.04176\,.$$

(The exact probability is $.0431\ldots$.)
B. Now W_6, the number of intervals until the sixth overload has a Negative Binomial$(6, p)$ distribution, and $W_6 = \sum_{j=1}^{6} Y_j$ where $Y_j \sim$ Geometric(p) are independent. Hence (again using the approximation from (a))

$$E(W_6) = 6\frac{1}{p} = 54.9\,,$$

and

$$Var(W_6) = 6\frac{q}{p^2} = 447.34\,.$$

(If we use the exact probability we get $E(W_6) = 6/(.1122) = 53.46$, and $Var(W_6) = 6(.8878)/(.1122)^2 = 423.14$.)

6. (32 points). Roll a pair of dice (one read and one white) with outcomes X_1 and X_2. Let $T = X_1 + X_2$, $D = X_1 - X_2$, and $M = \max\{X_1, X_2\}$. Evaluate:
(a) $P(D \geq 1 | X_1 + X_2 \leq 8)$.
(b) $P(M \leq 4 | T \geq 6)$.
(c) $P(|D| \leq 1 | M \geq 5)$.

Solution: The following table gives the values of T, D, and M for all the elementary outcomes:

x_2/x_1	1	2	3	4	5	6
6	(7,-5,6)	(8,-4,6)	(9,-3,6)	(10,-2,6)	(11,-1,6)	(12,0,6)
5	(6,-4,5)	(7,-3,5)	(8,-2,5)	(9,-1,5)	(10, 0,5)	(11,1,6)
4	(5,-3,4)	(6,-2,4)	(7,-1,4)	(8, 0,4)	(9, 1,5)	(10,2,6)
3	(4,-2,3)	(5,-1,3)	(6, 0,3)	(7, 1,4)	(8, 2,5)	(9,3,6)
2	(3,-1,2)	(4, 0,2)	(5, 1,3)	(6, 2,4)	(7, 3,5)	(8,4,6)
1	(2, 0,1)	(3, 1,2)	(4, 2,3)	(5, 3,4)	(6, 4,5)	(7,5,6)

(a) Thus we have

$$P(D \geq 1 | T \leq 8) = \frac{P(D \geq 1,\, T \leq 8)}{P(T \leq 8)} = \frac{11/36}{26/36} = \frac{11}{26}\,.$$

(b)

$$P(M \leq 4 | T \geq 6) = \frac{P(M \leq 4,\, T \geq 6)}{P(T \geq 6)} = \frac{6/36}{26/36} = \frac{6}{26}\,.$$

(c)

$$P(|D| \leq 1 | M \geq 5) = \frac{P(|D| \leq 1,\, M \geq 5)}{P(M \geq 5)} = \frac{6/36}{20/36} = \frac{6}{20}\,.$$

7. (24 points) Joe and Harry take turns rolling two dice with Joe going first. Joe wins if he rolls "6" or "8", while Harry wins if he rolls "4" or "7". Determine $P(\text{Joe wins})$ in this sequential contest.

Solution: First, in rolling two dice the probability of getting a sum of either 6 or 8 is $p_J \equiv 10/36$, while the probability of getting a sum of either 4 or 7 is $p_H \equiv 9/36$. Then

$$\begin{aligned} P(\text{Joe wins }) &= p_J + q_J q_H p_J + (q_J^2 q_H^2) p_J + (q_J^3 q_H^3) p_J + \cdots \\ &= p_J(1 + r + r^2 + \cdots) \qquad \text{where} \quad r \equiv q_J q_H \\ &= p_J \frac{1}{1-r} = p_J \frac{1}{1 - q_J q_H} \\ &= \frac{10/36}{1 - (26/36)(27/36)} = 0.6061\,. \end{aligned}$$

8. (24 points) Consider drawing (without replacement) from an urn consisting of 100 balls, 20 of which are labeled with the number 1, 30 of which are labeled with the number 5, 30 of which are labeled with the number 10, and 20 of which are labeled with the number 15. Determine the mean and standard deviation of:
(a) One randomly chosen ball.
(b) The sample mean of a sample of 40 randomly chosen (without replacement) balls.
(c) Use a central limit theorem to approximate $P(\overline{X}_n \geq 8.77)$.

Solution: (a) Now

$$\overline{a} = (20 \cdot 1 + 30 \cdot 5 + 30 \cdot 10 + 20 \cdot 15)/100 = 7.7\,,$$

while

$$\sigma_a^2 = (20 \cdot 1^2 + 30 \cdot 5^2 + 30 \cdot 10^2 + 20 \cdot 15^2)/100 = 8270/100 - (7.7)^2 = 23.41\,,$$

so that $\sigma_a = 4.838$. Hence for one draw from the urn we have

$$E(X_1) = \overline{a} = 7.7, \qquad Var(X_1) = \sigma_a^2 = 23.41\,,$$

and $\sqrt{Var(X_1)} = 4.838$.
(b) For the sample mean we have

$$E(\overline{X}_n) = \overline{a} = 7.7\,,$$

and

$$Var(\overline{X}_n) = \frac{\sigma_a^2}{n}\left(1 - \frac{n-1}{N-1}\right) = \frac{23.41}{40}\left(1 - \frac{39}{99}\right) \doteq .3547\,,$$

and hence $\sqrt{Var(\overline{X}_n)} \doteq .5956$.
(d) By the finite-sampling CLT,

$$\begin{aligned} P(\overline{X}_n \geq 8.77) &= P\left(\frac{\overline{X}_n - 7.7}{.5956} \geq \frac{8.77 - 7.7}{.5956}\right) \\ &\doteq P(Z \geq 1.7966) = 1 - .9641 = .0359\,. \end{aligned}$$

9. (24 points) Evaluate (or approximate) as accurately as you can, the probability of at least 7 honor cards in at least 12 of 200 hands of bridge. (Recall that a deck of 52 cards contains 16 “honor cards”, and that a bridge hand contains 13 cards.)

Solution: This goes in two stages: (a) first let $T_{13} = X_1 + \cdots + X_{13}$ be the total number of honor cards drawn (without replacement) from a deck of 52 cards. Thus $E(X_i) = 16/52 = 4/13$, $Var(X_i) = (4/13)(9/13)$, and the X_i's are dependent. Thus we have $E(T_{13}) = 13(4/13) = 4$ and $Var(T_{13}) = 13(4/13)(9/13)(1 - 12/51)$. Hence for one hand of bridge

$$\begin{aligned} P(T_{13} \geq 7) &= P(T_{13} \geq 6.5) \\ &= P\left(\frac{T_{13} - 4}{\sqrt{13(4/13)(9/13)(1-12/51)}} \geq \frac{6.5 - 4}{\sqrt{13(4/13)(9/13)(1-12/51)}}\right) \\ &\doteq P(Z \geq 1.7179) = 1 - .9571 \doteq .0429 \equiv p\,. \end{aligned}$$

The exact probability is

$$P(T_{13} \geq 7) = 1 - P(T_{13} \leq 6) = .04389\,.$$

(b) Now let $T_{200} = X_1 + \cdots + X_{200} \sim \text{Binomial}(200, p)$ be the number of bridge hands with 7 or more honor cards (in 200 bridge hands); here $p \doteq .0429$ from (a). Then, using the normal approximation from (a),

$$\begin{aligned} P(T_{200} \geq 12) &= P\left(\frac{T_{200} - 200(.0429)}{\sqrt{200(.0429)(1-.0429)}} \geq \frac{11.5 - 200(.0429)}{\sqrt{200(.0429)(1-.0429)}}\right) \\ &\doteq P(Z \geq 1.018968) \doteq 1 - .8461 = .1539\,. \end{aligned}$$

If we calculate using the exact value in part (a), but use the CLT to approximate in the second stage again, we get

$$\begin{aligned} P(T_{200} \geq 12) &= P\left(\frac{T_{200} - 200(.04389)}{\sqrt{200(.04389)(1-.04389)}} \geq \frac{11.5 - 200(.04389)}{\sqrt{200(.04389)(1-.04389)}}\right) \\ &\doteq P(Z \geq .9396) \doteq 1 - .8264 = .1736\,. \end{aligned}$$

The exact probability is

$$P(Binomial(200, .04389) \geq 12) = 1 - P(Binomial(200, .04389) \leq 11) = 1 - .8286 = .1714\,.$$

Using the normal approximation from (a), and a Poisson Approximation in stage (b) instead of the normal approximation yields

$$\begin{aligned} P(T_{200} \geq 12) &\doteq P(X \geq 12) \qquad X \sim \text{Poisson}(\lambda = 200(.0429) = 8.58) \\ &= 1 - P(X \leq 11) = .1582\,. \end{aligned}$$

Using the exact probability from part (a), and a Poisson approximation in stage (b), we get

$$\begin{aligned} P(T_{200} \geq 12) &\doteq P(X \geq 12) \qquad X \sim \text{Poisson}(\lambda = 200(.04389) = 8.778) \\ &= 1 - P(X \leq 11) = .1762\,. \end{aligned}$$