PROBABILITY AND STATISTICS 550.310 SPRING 2009
EXAM #2

Name: Bruno Jedynak **Section:**

No document. No calculator. The exam is worth 110 points.

1. Multiple Choices (20 points)

Clearly circle the letter of the most correct answer to each question.

(1) For any random variables X and Y, , $V[X+Y] \geq V[X] + V[Y]$
  (a) true
  (b) false

(2) For any random variables X and Y, if X and Y are independent then $E(XY) = E(X)E(Y)$
  (a) true
  (b) false

(3) For any random variable X, $E(X^2) \geq E^2(X)$.
  (a) true
  (b) false $V[X] = E[X^2] - E[X]^2 \geq 0$

(4) If X and Y are independent and both Uniform over the interval $[0, 1]$ then $X + Y$ is Uniform over the Interval $[0, 2]$
  (a) true
  (b) false

$V[X+Y] = V[X] + V[Y] + 2\,\mathrm{Cov}(X,Y)$

In general, $\mathrm{Cov}(X,Y)$ can take any value.

If X and Y are independent then $\mathrm{Cov}(X,Y) = 0$

and $V[X+Y] = V[X] + V[Y]$

Date: February 20th, 2009.

2. Problems

Show all work to obtain credits. For each single question, simplify until obtaining an irreducible fraction.

2.1. **joint Pdf.** The joint probability density function of X and Y is given by

$$f(x,y) = \frac{6}{7}(x^2 + \frac{xy}{2}), 0 \leq x \leq 1, 0 \leq y \leq 2$$

(1) Verify that this is indeed a joint density function.

$$\int_{x=0}^{1}\int_{y=0}^{2} \frac{6}{7}\left(x^2+\frac{xy}{2}\right) dy\, dx = \int_{x=0}^{1} \frac{6}{7}\left(x^2 y + \frac{x}{2}\frac{y^2}{2}\right)\Bigg|_{y=0}^{2} dx$$

$$= \int_{x=0}^{1} \frac{6}{7}\left(2x^2+x\right) dx = \frac{6}{7}\left(\frac{2}{3}x^3+\frac{1}{2}x^2\right)\Bigg|_{x=0}^{1} = \frac{6}{7}\left(\frac{2}{3}+\frac{1}{2}\right)$$

$$= \frac{6}{7}\left(\frac{7}{6}\right) = 1$$

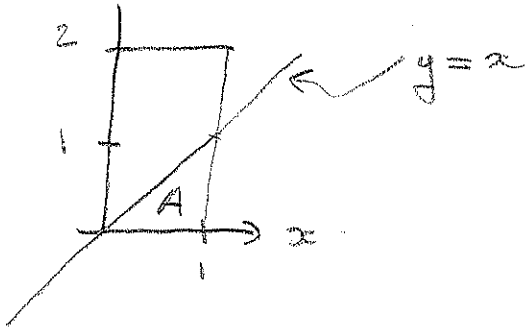
(2) Compute the marginal density of X.

Hint: do not forget to indicate for which values of x the density vanishes.

$$0 \leq x \leq 1, \quad f_X(x) = \int_{y=0}^{2} \frac{6}{7}\left(x^2+\frac{xy}{2}\right) dy$$

$$f_X(x) = \frac{6}{7}\left(2x^2+x\right), \qquad 0 \leq x \leq 1$$

(3) Compute $P(X > Y)$



$$P(X>Y) = \iint_A f(x,y)\,dx\,dy$$

$$= \int_{x=0}^{1} \int_{y=0}^{x} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) dy\, dx$$

$$= \int_{x=0}^{1} \frac{6}{7}\left(x^2 y + \frac{xy^2}{4}\right)\Big|_{y=0}^{x} dx$$

$$= \int_{x=0}^{1} \frac{6}{7}\left(x^3 + \frac{x^3}{4}\right) dx = \frac{\cancel{6}^{\,3}}{7}\,\frac{5}{\cancel{4}_{\,2}}\,\frac{1}{4} = \frac{15}{56}$$

2.2. **Playing big once, or playing little many times.** We consider the following game: The player can choose how much he plays. When playing \$a, his net gain, notated X, is Normally distributed with mean \$a and standard deviation \$a. Hence $X \sim N(a, a^2)$.

We give the following values for $\phi(z) = P(Z < z)$ when Z is Standard Normal, i.e $Z \sim N(0,1)$.

z	0	1	2	3	4	5	6	7	8	9	10
$\phi(z)$	0.5	0.84	0.98	0.998	0.9999683	0.9999997	1	1	1	1	1

(1) The player plays once, for an amount of $a = \$100$. What is the probability that he ends up with less money that he started with, i.e $X < 0$.

$$X \sim N(100; 10{,}000)$$

$$P(X<0) = P\left(\frac{X-100}{\sqrt{10{,}000}} < \frac{0-100}{\sqrt{10{,}000}}\right)$$

$$= P(Z<-1), \quad Z \sim N(0,1)$$

$$= 1 - P(Z<1)$$

$$= 1 - .84 = .16$$

(2) The player plays 100 games. During each game i, $1 \leq i \leq 100$, he plays the same amount $a = \$1$ and his gain is X_i. We assume that the gains $X_1, \ldots, X_{100}$ are independent. His total net gain is $S = X_1 + \ldots, +X_{100}$. What is the probability that he ends up with less money than he started with, i.e $S < 0$.

$$E[S] = 100; \quad V[S] = 100; \quad S \sim N(100; 100)$$

$$P(S<0) = P\left(\frac{S-100}{\sqrt{100}} < \frac{0-100}{\sqrt{100}}\right)$$

$$= P(Z < -10); \quad Z \sim N(0;1)$$

$$= 1 - P(Z < 10) \simeq 0$$

2.3. **Basketball.** A club basketball team will play a 64-game season. Thirty two of these games are against class A teams and 32 are against class B teams. The outcomes of all the games are independent. The team will win each game against a class A opponent with probability .5 and it will win each game against a class B team with probability .75. Let X denote its total number of victories in the season.

(1) Let X_A and X_B represent respectively the number of victories against class A and class B teams. What are the distributions of X_A and X_B? specify all the parameters.

$X_A \sim \text{Binomial}(32; .5)$

$X_B \sim \text{Binomial}(32; .75)$

(2) Give, approximatively, the probability that the team wins more than 44 games during the season. Justify precisely your answer in order to obtain partial credits.

Hint: Use the central limit theorem, use the approximation $44 \sim 40 + \sqrt{14}$, and the table for ϕ given in the previous problem.

Let $X = X_A + X_B$

$E[X_A] = 32(.5) = 16$; $V[X_A] = 32(.5)(.5) = 8$

$E[X_B] = 32(.75) = 24$; $V[X_4] = 32(.75)(.25) = 6$

Using the C.L.T, $X_A \approx N(16; 8)$; $X_B \approx N(24; 6)$

Since X_A and X_B are independent,

$X_A + X_B \approx N(40, 14)$

$P(X > 44) = P\left(\frac{X-40}{\sqrt{14}} > \frac{44-40}{\sqrt{14}}\right)$

$\approx P(Z > 1) = .16$

3. Discrete distributions

The random variable is denoted by the letter X. The range is denoted by R_X. The point mass function (pmf) is denoted by the letter f. It is given for $x \in R_X$. The cumulative distribution function (cdf) is denoted by the letter F. It is given if it is of special interest. The expected value is $E(X)$, the variance is $V(X)$.

(1) **Discrete Uniform:** X is Discrete uniform over the range

$$R_X = \{a, a+1, \ldots, b\}$$

$a < b$

$$f(x) = \frac{1}{b-a+1}$$

$$E(X) = \frac{a+b}{2}, V(X) = \frac{(b-a+1)^2 - 1}{12}$$

(2) **Binomial and Bernouilli:** X is Binomial(n,p)

$$R_X = \{0, 1, \ldots, n\}$$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np, V(X) = np(1-p)$$

The Bernouilli(p) distribution is Binomial(1,p).

(3) **Hypergeometric:** If X is the number of S's in a completely random sample of size n drawn without replacement from a population consisting of M S's and N-M F's, the distribution of X is hypergeometric.

$$R_X = \{\max(0, n-N+M), \ldots, \min(n, M)\}$$

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$E(X) = n\frac{M}{N}, V(X) = \frac{N-n}{N-1} n \frac{M}{N}(1 - \frac{M}{N})$$

(4) **Negative Binomial and Geometric:** Suppose that a sequence of independent trials, each with probability of success p, is performed until there are r success in all. X is the number of failures that precede the rth success.

$$R_X = \{0, 1, \ldots\}$$

$$f(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

$$E(x) = \frac{r(1-p)}{p}, V(X) = r(1-p)/p^2$$

When $r = 1$ this distribution is also called the Geometric distribution with parameter p.

(5) **Poisson:** X is Poisson(λ).

$$R_X = \{0, 1, \ldots\}$$

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

$$E(X) = V(X) = \lambda$$

4. Continuous distributions

The random variable is denoted by the letter X. The range is denoted by R_X. The probability density function (pdf) is denoted by the letter f. It is given for $x \in R_X$. The cumulative distribution function (cdf) is denoted by the letter F. It is given for $t \in R_X$ if it is of special interest. The expected value is $E(X)$, the variance is $V(X)$.

(1) **Continuous Uniform** X is Uniform([a,b]), with $a < b$.
$$R_X = [a, b]$$
$$f(x) = \frac{1}{b-a}$$
$$F(t) = \frac{t-a}{b-a}$$
$$E(X) = \frac{a+b}{2}, V(X) = \frac{(b-a)^2}{12}$$

(2) **Normal Distribution** X is Normal(μ,σ^2).
$$R_X = (-\infty, +\infty)$$
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$
$$E(X) = \mu, V(X) = \sigma^2$$
The Normal(0,1) is called the Standart Normal. Its cdf is denoted ϕ. It verifies
$$\phi(-t) = 1 - \phi(t)$$

(3) **Gamma** X is Gamma(α, β). $\alpha > 0$ $\beta > 0$.
$$R_X = [0, +\infty)$$
$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$
with
$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$
$$E(X) = \alpha\beta, V(X) = \alpha\beta^2$$

(4) **Exponential** The Exponential distribution with parameter λ is Gamma($\alpha = 1, \beta = \frac{1}{\lambda}$)

(5) **Chi-square** The Chi-square distribution with n degres of freedom is Gamma($\alpha = \frac{n}{2}, \beta = 2$)

(6) **Beta** X is *Beta*(α, β). $\alpha > 0$, $\beta > 0$.
$$R_X = [0; 1]$$
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$
$$E(X) = \frac{\alpha}{\alpha+\beta}, V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

(7) **Student** X is Student(n) $n > 0$,

$$R_X = \mathbb{R}$$

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$$E(X) = 0, V(X) = \frac{n}{n-2} \text{ when } n > 2$$

(8) **LogNormal** X is LogNormal with parameter μ and σ^2.

$$R_X = (0, +\infty)$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right)$$

$$E(X) = \exp(\mu + \frac{\sigma^2}{2}), V(X) = \exp(2\mu + \sigma^2)\left(\exp(\sigma^2) - 1\right)$$

5. Variance, Covariance and Correlation

The Variance of a random variable X is

$$V[X] = E[(X - E[X])^2] = E[X^2] - E^2[X]$$

And

$$V[aX + b] = a^2V[X]$$

The covariance of 2 random variables is

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

$$Cov(aX + b, cY + d) = acCov(X,Y)$$

$$Cov(X,X) = V[X]$$

6. Linear Combinations of random Variables

Let $X_1, \dots, X_n$ be n random variables and $a_1, \dots, a_n$ be n real numbers. Linearity of the Expected value:

$$E[\sum_{i=1}^{n} a_iX_i] = \sum_{i=1}^{n} a_iE[X_i]$$

The covariance is a bilinear form:

$$Cov(\sum_{i=1}^{n} a_iX_i, \sum_{i=1}^{n} b_iY_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ib_jCov(X_i, Y_j)$$

The variance is a Quadratic form:

$$V[\sum_{i=1}^{n} a_iX_i] = Cov(\sum_{i=1}^{n} a_iX_i, \sum_{i=1}^{n} a_iX_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i, X_j)$$

7. Random sample

$X_1, \ldots, X_n$ is a random sample when $X_1, \ldots, X_n$ are independent and have the same distribution.

In this case

$$E[X_1] = \ldots = E[X_n] \text{ and } V[X_1] = \ldots = V[X_n]$$

and

$$Cov(X_i, X_j) = 0 \text{ for } i \neq j$$

hence

$$V[X_1 + \cdots + X_n] = V[X_1] + \ldots + V[X_n]$$

Law of Large Numbers: for any function g such that $E[g(X_1)]$ exists,

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} g(X_i) = E[g(X_1)]$$

Central Limit Theorem:

Let

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

if $V[X_1]$ exists, the distribution of $\frac{\sqrt{n}}{\sigma_{X_1}}(\bar{X}_n - E[X_1])$ converges, when $n \to +\infty$ to the Standard Normal distribution.