MATH/STAT 3360, Probability
FALL 2011
Toby Kenney
Sample Final Examination

This Sample Examination has more questions than the actual exam, in order to cover a wider range of questions.

1. *(a) A truncated standard normal random variable has probability density function*

$$f_X(x) = \begin{cases} \frac{e^{-\frac{x^2}{2}}}{\int_a^b e^{-\frac{x^2}{2}}} & \text{if } a \leqslant x \leqslant b \\ 0 & \text{otherwise} \end{cases}$$

That is, it is a normal distribution conditional on lying in the interval $[a, b]$.

What is the expectation of a truncated standard normal distribution?

The expectation of a truncated standard normal is given by $E(T) = \frac{\int_a^b xe^{-\frac{x^2}{2}}\,dx}{\int_a^b e^{-\frac{x^2}{2}}\,dx} = [\frac{-e^{-\frac{x^2}{2}}}{\sqrt{2\pi}(\Phi(b)-\Phi(a))}]_a^b = \frac{e^{-\frac{a^2}{2}}-e^{-\frac{b^2}{2}}}{\sqrt{2\pi}(\Phi(b)-\Phi(a))}$.

(b) An investment advisor tells you that the amount by which the market will increase next year is normally distributed with mean 10% and standard deviation 10%. The bank offers an investment product which pays the value of the market, minus a 2% fee if the market increases by between 2% and 32%. If the market rises by less than 2%, then it pays out your original investment; if the market rises by more than 32%, it pays out your original investment plus 30%. What is the expected value of this investment? [Hint: divide into the three cases where the market increases by less than 2%, the market increases by between 2% and 32%, and the market increases by over 32%.]

The probability that the markest increases by less than 2% is $\Phi\left(\frac{2-10}{10}\right) = \Phi(-0.8) = 1-\Phi(0.8)$. The probability that the market increases by more than 32% is $1-\Phi\left(\frac{32-10}{10}\right) = 1-\Phi(2.2)$. Therefore, we can calculated the expected gain by conditioning on whether the market increases by more than 32%, 2–32% or less than 2%. Let X be the percentage by which the market increases. We then have the following:

X	Probability	Expected value
$X < 2$	$1-\Phi(0.8)$	0
$2 \leqslant X \leqslant 32$	$\Phi(2.2)+\Phi(0.8)-1$	$8+10\frac{e^{-\frac{0.8^2}{2}}-e^{-\frac{2.2^2}{2}}}{\sqrt{2\pi}(\Phi(2.2)+\Phi(0.8)-1)}$
$X > 32$	$1-\Phi(2.2)$	30

Therefore, the expected increase of the investment is $8(\Phi(2.2)+\Phi(0.8)-1)+10(\frac{e^{-0.32}-e^{-2.42}}{\sqrt{2\pi}})+30(1-\Phi(2.2)) = 8(0.9861+0.7881-1)+10\times 0.254217+30\times 0.0139 = 9.15\%$.

2. *If X is uniformly distributed on [-1,4], and Y is uniformly distributed on the interval $[0,3]$, find the moment generating function of $X - Y$.*

   The moment generating function of X is $M_X(t) = \frac{e^{4t}-e^{-t}}{5t}$, and the moment generating function of $-Y$ is $M_{-Y}(t) = \frac{1-e^{-3t}}{3t}$. Therefore, the moment generating function of $X-Y$ is $\left(\frac{e^{4t}-e^{-t}}{5t}\right)\left(\frac{1-e^{-3t}}{3t}\right) = \frac{e^{4t}-e^{t}-e^{-t}+e^{-4t}}{15t^2}$.

3. *An investor's annual profit has an expectation of \$2,000 and a variance of 100,000,000. (Profits in different years are independent and identically distributed.)*

   (a) After n years, for large n, what is the approximate distribution of the investors average annual profit?

   The approximate distribution is normal with mean 2000 and variance $\frac{100000000}{n}$.

   (b) After how many years is the probability that the investor has made an overall loss during those years less than 0.001?

   The investor has made an overall loss if his average annual profit is less than 0. The probability of this is $1 - \Phi\left(\frac{2000}{\sqrt{\frac{100000000}{n}}}\right) = 1 - \Phi\left(\frac{\sqrt{n}}{5}\right)$. Now $\Phi(3.09) = 0.999$, so the probability is less than 0.001 when $\frac{\sqrt{n}}{5} > 3.09$, which happens when $\sqrt{n} > 15.45$, or $n > 238.7025$, so after 239 years.

   (c) If at any point, the investor has made a total loss of over \$100,000, he must pay a fine. After how many years is the danger that he will have to pay this fine greatest, and what is the danger after this many years [assuming the approximation from (a) is valid]?

   He has to pay the fine after n years if the average annual loss is over $\frac{100000}{n}$. The probability of this is $1 - \Phi\left(\frac{2000+100000n^{-1}}{10000n^{-\frac{1}{2}}}\right)$. This is maximised when $\frac{2000+100000n^{-1}}{10000n^{-\frac{1}{2}}}$ is minimised. Now $\frac{2000+100000n^{-1}}{10000n^{-\frac{1}{2}}} = \frac{\sqrt{n}}{5} + \frac{10}{\sqrt{n}}$, which is minimised when $\frac{\sqrt{n}}{5} = \frac{10}{\sqrt{n}}$, or when $n = 50$. After 50 years, the risk is $1 - \Phi\left(\frac{4000\sqrt{50}}{10000}\right) = 1 - \Phi(2.828) = 1 - 0.9976 = 0.0024$.

   (d) Using the one-sided Chebyshev inequality, what is the largest probability that he will have to pay the fine after n years.

   After n years, the total gain X has expectation $2000n$ and variance $100000000n$, so the probability that $X < -100000$ is at most $\frac{100000000n}{100000000n+(2000n+100000)^2}$.

   (e) For what number of years is this maximised, and what is the upper bound on the probability in this case?

   This is maximised when $\frac{(2000n+100000)^2}{100000000n}$ is minimised, or equivalently, when $\frac{2000n+100000}{10000\sqrt{n}}$ is minimised. This happens when $2000n = 100000$, or equivalently when $n = 50$. In this case, the upper bound on the probability is $\frac{5000000000}{5000000000+200000^2} = \frac{50}{50+20^2} = \frac{5}{45} = \frac{1}{9}$.

4. *In a class of 100 students, the professor wants to determine what proportion of the students can answer a simple calculus question. The professor decides to test a random sample of 10 students. In fact 35 of the students could answer the question.*

   (a) What is the expected number of students sampled who answer the question correctly?

   Each student has probability 0.35 of answering correctly, so the expected number who answer correctly is $10 \times 0.35 = 3.5$.

   (b) What is the variance of the number of correct answers if the students are sampled with replacement, i.e. one student could be tested more than once?

   In this case, the number of correct answers is binomial, so the variance is $10 \times 0.35 \times 0.65 = 2.275$.

   (c) What is the variance of the number of correct answers if the 10 students sampled must be 10 different students?

   Let X_i be the event that the ith student answers correctly. Then $P(X_i X_j = 1) = \frac{\binom{35}{2}}{\binom{100}{2}} = \frac{35\times34}{100\times99}$. Therefore, if X is the number of students who answer correctly, then $E(X(X-1)) = \sum_{i\neq j} \frac{35\times34}{100\times99} = \frac{10\times9\times35\times34}{100\times99} = \frac{7\times17}{11} = 10.1818$. We have that $\mathrm{Var}(X) = E(X(X-1)) + E(X) - (E(X))^2 = 10.1818 + 3.5 - 3.5^2 = 1.4318$.

5. *If X is exponentially distributed with parameter λ_1 and Y is exponentially distributed with parameter λ_2, what is the probability density function of $X+Y$? [Assume $\lambda_1 \neq \lambda_2$.]*

   We know that $f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(y) f_Y(x-y)dy = \int_0^x \lambda_1\lambda_2 e^{-\lambda_1 y} e^{-\lambda_2(x-y)}dy = \lambda_1\lambda_2 e^{-\lambda_2 x}\int_0^x e^{(\lambda_2-\lambda_1)y}dy$. Now $\int_0^x e^{(\lambda_2-\lambda_1)y}dy = \left[\frac{e^{(\lambda_2-\lambda_1)y}}{(\lambda_2-\lambda_1)}\right]_0^x = \frac{e^{(\lambda_2-\lambda_1)x}-1}{(\lambda_2-\lambda_1)}$. Therefore, we get $f_{X+Y}(x) = \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}(e^{-\lambda_1 x} - e^{-\lambda_2 x})$.

6. *X is normally distributed with mean 3 and standard deviation 3 and Y is normally distributed with mean 7 and standard deviation 4. What is $P(Y > X)$?*

   Since X is normally distributed with mean 3 and standard deviation 3, $-X$ is normally distributed with mean -3 and standard deviation 3, so $Y-X$ is a sum of two normal random variables, so is normally distributed with mean 4 and variance $3^2+4^2 = 25$. Therefore $P(Y > X) = P(Y-X > 0) = \Phi\left(\frac{4}{\sqrt{25}}\right) = \Phi(0.8) = 0.7881$.

7. *Let X have an exponential distribution with parameter λ. Suppose that given $X = x$, Y is normally distributed with mean x and variance σ^2.*

   (a) What is the joint density function of X and Y?

   The joint density function of X and Y is $f_{X,Y}(x,y) = \frac{\lambda}{\sqrt{2\pi}\sigma} e^{-\lambda x} e^{-\frac{(y-x)^2}{2\sigma^2}}$.

(b) What is the covariance of X and Y?

We have that $E(Y|X=x)=x$, so $E(XY|X=x)=x^2$. Now $E(XY)=E(E(XY|X))=E(X^2)=(E(X))^2+\mathrm{Var}(X)=18$. Similarly, $E(Y)=E(E(Y|X))=E(X)=3$, so $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=18-3\times 3=9$.

(c) Conditional on $Y=3$, what is the density function of X? [You do not need to calculate the constant factor.]

The condition density of X is $f_{X|Y}(x|3)=ke^{-\lambda x}e^{-\frac{(x-3)^2}{2\sigma^2}}$ for some constant k, for $x\geqslant 0$, and 0 for $x<0$.

[By completing the square, we see that $e^{-\lambda x}e^{-\frac{(x-3)^2}{2\sigma^2}}=e^{-\frac{x^2-(6-2\lambda\sigma^2)x+9}{2\sigma^2}}$. so X has a truncated normal distribution (i.e. a normal distribution conditional on $X>0$, with mean $3-\lambda\sigma^2$ and variance σ.]

8. *On average, in a certain course, 20% of students get an 'A' and 15% students get a 'B'. There are 25 students taking the course one year. What is the probability that 5 students get an 'A' and 5 students get a 'B'?*

   This is a multinomial distribution. The probability that 5 students get an 'A' and 5 students get a 'B' is therefore $\binom{25}{5,5,15}0.2^5 0.15^5 0.65^{15}=0.0313$.

9. *(a) How many different ways are there to distribute numbers 1–10 among 20 people in such a way that each number is used exactly twice? [Hint: order the people in any way, then the different ways correspond to sequences of the numbers 1–10 such that each number occurs twice.]*

   This is the number of different rearrangements of the "word" "12345678910123456789 10", which is $\frac{20!}{2!2!2!2!2!2!2!2!2!2!}$.

   (b) How many ways are there to divide 20 people into 10 pairs if the order of the pairs doesn't matter? [Hint: how many ways are there to divide 20 people into 10 pairs if the pairs are numbered 1–10?]

   If the pairs are numbered 1–10, then the number of ways is $\frac{20!}{2!2!2!2!2!2!2!2!2!2!}$. The same effect can be achieved by dividing the people into 10 pairs first, then numbering the pairs in any of 10! ways. Therefore, the number of ways to divide the people into 10 pairs is $\frac{20!}{10!2!2!2!2!2!2!2!2!2!2!}$.

10. *The number of claims an insurance company needs to pay follows a Poisson distribution with parameter 4.5. What is the probability that the company needs to pay out 5 or more claims?*

    Let C be the number of claims paid out. Now $P(C\geqslant 5)=1-P(C=0)-P(C=1)-P(C=2)-P(C=3)-P(C=4)=1-e^{-4.5}\left(1+4.5+\frac{4.5^2}{2}+\frac{4.5^3}{6}+\frac{4.5^4}{24}\right)=0.468$.

11. *The number of visitors to a particular website on a given day is approximately normally distributed with mean 12000 and variance 2000^2. A company is considering placing an advertisement on this website. It predicts*

that each visitor to the website will order its product with probability 0.02, and that all visitors to the website act independently.

(a) What is the expectation and variance of the number of orders the company receives?

Let X be the number of visitors to the website and Y the number of orders the company receives. We have that $E(Y|X = x) = 0.02x$, $E(Y^2|X = x) = 0.0004x^2 + 0.0196x$. Therefore, $E(Y) = E(E(Y|X)) = E(0.02X) = 0.02E(X) = 240$, and $E(Y^2) = E(E(Y^2|X)) = E(0.0004X^2 + 0.0196X) = 0.0004E(X^2) + 0.0196E(X) = 0.0004(12000^2 + 2000^2) + 0.0196 \times 12000 = 59200 + 235.2 = 59435.2$. Therefore $\mathrm{Var}(Y) = E(Y^2) - (E(Y))^2 = 59435.2 - 57600 = 1835.2$.

(b) Using a normal approximation, what is the probability that the number of orders the company receives is more than 300?

The probability that the company receives more than 300 orders is $1 - \Phi\left(\frac{300.5-240}{\sqrt{1835.2}}\right) = 1 - \Phi(1.41) = 1 - 0.9207 = 0.0793$.

(c) If 300 is the maximum number of orders that the company is able to satisfy (that is, if the company receives over 300 orders, it sells only 300 products) what is the expected number of products sold?

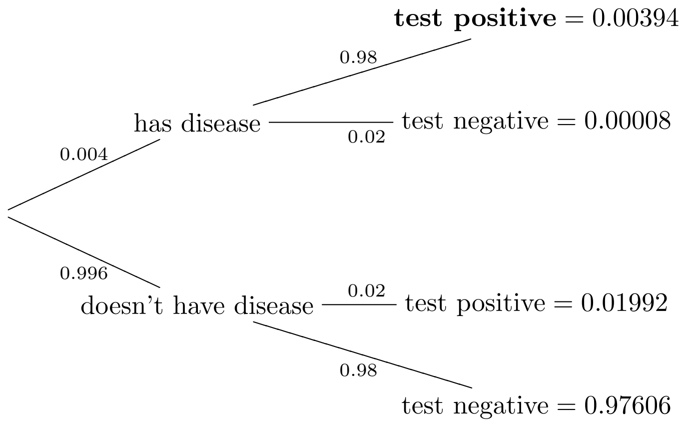
$\int_{-\infty}^{a} xe^{-\frac{x^2}{2}} dx = [-e^{-\frac{x^2}{2}}]_{-\infty}^{a} = -e^{-\frac{a^2}{2}}$.

So the expectation of a truncated standard normal is given by $E(T) = -\frac{e^{-\frac{a^2}{2}}}{\sqrt{2\pi}} + a(1 - \Phi(a))$.

In particular, when $a = 1.41$, then we get $E(T) = -\frac{e^{-\frac{a^2}{2}}}{\sqrt{2\pi}} + a(1 - \Phi(a)) = 1.41(0.0793) - 0.147 = -0.035$.

Therefore, the expectation of Y is given by $E(Y) = \sqrt{1835.2}E(T) + 240 = 238.49$.

12. *A patient is given a routine test for a rare disease. The disease affects 4 people in 1000. The test is 98% accurate, so there is a 2% chance of giving the wrong result. The test result is positive (i.e. indicates the patient has the disease). What is the probability that the patient actually has the disease?*



So the total probability of a positive test result is $0.00394 + 0.01992 = 0.002386$, and the probability of having the disease given a positive test result is $\frac{0.00394}{0.02386} = 0.165$.

13. *A gambler starts with \$2, and continues to make a series of \$1 bets with probability $\frac{1}{2}$ until he loses all his money, or until he has \$5. What is the expected number of bets until he has lost all his money or reached \$5? [Hint: let e_n be the expected number of bets conditional on having \$$n$, and condition on the outcome of the next bet to get a relation between the e_n.*

Let e_n be the expected time for him to lose all his money or reach \$5, given that he currently has \$$n$. Now clearly, since his probability of winning and losing are both $\frac{1}{2}$, we have that $e_n = 1 + \frac{1}{2}(e_{n+1} + e_{n-1})$. Furthermore, we have that $e_0 = 0$ and $e_5 = 0$. This gives the equations:

$$
\begin{aligned}
e_1 &= 1 + \frac{1}{2}(e_2 + 0) \\
e_2 &= 1 + \frac{1}{2}(e_3 + e_1) \\
e_3 &= 1 + \frac{1}{2}(e_4 + e_2) \\
e_4 &= 1 + \frac{1}{2}(0 + e_3)
\end{aligned}
$$

Rearranging these, we get:

$$
\begin{aligned}
e_2 &= 2(e_1 - 1) \\
e_3 &= 2(e_2 - 1) - e_1 = 3e_1 - 6 \\
e_4 &= 2(e_3 - 1) - e_2 = 4e_1 - 12
\end{aligned}
$$

Now the last equation reads $4e_1-12 = 1+\frac{1}{2}(3e_1-6)$, which gives $5e_1 = 20$, or $e_1 = 4$, and therefore we get $e_2 = 6$, $e_3 = 6$, and $e_4 = 4$, so the expected time until he finishes is 4 bets.

14. *Random variables X and Y have joint density function $f_{X,Y}(x,y) = \frac{1}{\sqrt{10\pi}}e^{-\frac{5x^2+5y^2-6xy}{10}}$.*

(a) What is the joint density function of $Z = X + Y$ and $W = X - Y$?

$$\frac{\partial z}{\partial x} = 1 \qquad \frac{\partial z}{\partial y} = 1$$
$$\frac{\partial w}{\partial x} = 1 \qquad \frac{\partial w}{\partial y} = -1$$

So we get that $dxdy = 2dzdw$. Therefore, the joint density function of (Z,W) is given by $f_{Z,W}(z,w) = \frac{2}{\sqrt{10\pi}}e^{-\frac{5x^2+5y^2-6xy}{10}}$. Now $x = \frac{1}{2}(z+w)$ and $y = \frac{1}{2}(z-w)$, so $5x^2+5y^2-6xy = \frac{5}{4}(2z^2+2w^2) - \frac{6}{4}(z^2-w^2) = z^2+4w^2$, so $f_{Z,W}(z,w) = \frac{2}{\sqrt{10\pi}}e^{-\frac{z^2+4w^2}{10}}$.

(b) Are W and Z independent?

$f_{Z,W}(z,w) = \frac{2}{\sqrt{10\pi}}e^{-\frac{z^2+4w^2}{10}} = \frac{2}{\sqrt{10\pi}}e^{-\frac{z^2}{10}}e^{-\frac{4w^2}{10}}$ is a product of a function of z and a function of w, so Z and W are independent.

(c) What is $\mathrm{Cov}(X,Y)$*? [Hint: consider* $\mathrm{Var}(Z) - \mathrm{Var}(W)$*.]*

Recall that $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)+2\,\mathrm{Cov}(X,Y)$, and $\mathrm{Var}(X-Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)-2\,\mathrm{Cov}(X,Y)$. Therefore, $\mathrm{Var}(X+Y)-\mathrm{Var}(X-Y) = 4\,\mathrm{Cov}(X,Y)$. Now from (a) and (b), we see that $\mathrm{Var}(X+Y) = \mathrm{Var}(Z) = 5$, and $\mathrm{Var}(X-Y) = \mathrm{Var}(W) = 1.25$, so $4\,\mathrm{Cov}(X,Y) = 5-1.25 = 3.75$, so $\mathrm{Cov}(X,Y) = 0.9375$.

15. *Let X and Y are independent normal random variables with mean 0 and variances σ_1^2 and σ_2^2 respectively.*

(a) What is the joint density function of X and Y?

$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{\sigma_2^2x^2+\sigma_1^2y^2}{2(\sigma_1^2\sigma_2^2)}}$.

(b) If (R,Θ) are the polar coordinates of (X,Y) (that is, $R = \sqrt{X^2+Y^2}$ and Θ is the solution to $R\cos\Theta = X$ and $R\sin\Theta = Y$) then what is the joint density function of R and Θ?

If $x = r\cos\theta$ and $y = r\sin\theta$, then

$$\frac{\partial x}{\partial r} = \cos\theta$$
$$\frac{\partial x}{\partial \theta} = -r\sin\theta$$
$$\frac{\partial y}{\partial r} = \sin\theta$$
$$\frac{\partial y}{\partial \theta} = r\cos\theta$$

Therefore, we get that $dxdy = rdrd\theta$, so that the joint density function of (R, Θ) is $f_{R,\Theta}(r,\theta) = \frac{r}{2\pi\sigma_1\sigma_2}e^{-\frac{\sigma_2^2x^2+\sigma_1^2y^2}{2(\sigma_1^2\sigma_2^2)}} = \frac{r}{2\pi\sigma_1\sigma_2}e^{-\frac{r^2(\sigma_2^2\cos^2\theta+\sigma_1^2\sin^2\theta)}{2(\sigma_1^2\sigma_2^2)}}$.

(c) Calculate the marginal distribution of Θ.

The marginal density of Θ is given by integrating R, so $f_\Theta(\theta) = \int_0^\infty \frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{r^2(\sigma_2^2\cos^2\theta+\sigma_1^2\sin^2\theta)}{2(\sigma_1^2\sigma_2^2)}} = \frac{r}{2\pi\sigma_1\sigma_2}\left[-\frac{e^{-\frac{r^2(\sigma_2^2\cos^2\theta+\sigma_1^2\sin^2\theta)}{2\sigma_1^2\sigma_2^2}}}{\frac{(\sigma_2^2\cos^2\theta+\sigma_1^2\sin^2\theta)}{\sigma_1^2\sigma_2^2}}\right]_0^\infty = \frac{\sigma_1\sigma_2}{2\pi(\sigma_2^2\cos\theta+\sigma_1^2\sin\theta)}$.