# Midterm Examination Solutions, Statistics 220, Fall 2005

19 October 2005

## 1 Binomial vs. Poisson

$X$ is a binomial random variable, $n = 10$, $p = 0.1$. $Y$ is a Poisson random variable, $\lambda = 1$.
(a) What is $\mathbf{E}[X]$? (1 pt) What is $\mathrm{Var}(X)$? (1 pt) What is the probability that $X = 2$? (3 pts)
(b) What is $\mathbf{E}[Y]$? (1 pt) What is $\mathrm{Var}(Y)$? (1 pt) What is the probability that $Y = 2$? (5 pts)

(a) The expectation of a binomial is $np = 1$. The variance is $np(1-p) = 0.9$.

$$\begin{aligned}\Pr(X=2) &= \binom{10}{2}(0.1)^2(0.9)^8 \\ &= \frac{10\cdot 9\cdot 8!}{2\cdot 8!}(0.1)^2(0.9)^8 = 45(0.1)^2(0.9)^8 = 0.1937102\end{aligned}$$

(b) The expectation of a Poisson is $\lambda = 1$. The variance is $\lambda = 1$.

$$\Pr(()\,Y = 2) = e^{-1}\frac{1^2}{2!} = 0.1839397$$

## 2 Camels

Bactrian camels (*Camelus bactrianus*) are used as beasts of burden in Central Asia. A survey of the camels available in a certain town in Afghanistan shows that their mean weight is 1600 pounds, with a standard deviation of 150 pounds.
(a) What are the mean and standard deviation of the camels' weights in kilograms? (2 pts) *Hint*: 1 kg = 2.2 lb.
(b) The camels are now loaded with water casks weighing exactly 120 kilograms each. What is the mean and standard deviation of their total weight? (3 pts)

(a) Multiplying the measurements by a constant multiplies the mean and the standard deviation by the same constant. So the mean weight is 727 kg, and the standard deviation of the weight is 68 kg.

(b) Adding a constant increases the mean by that amount, but doesn't change the standard deviation. Mean loaded weight, 847 kg, standard deviation 68 kg.

# 3 Coins

(a) What are the probabilities of getting the following sequences of heads and tails from 10 consecutive tosses of a fair coin? (2 pts)
(i) HHHHHHHHHH
(ii) HTHTHTHTHT
(iii) HHTTTTTTHH
(b) What is the probability of getting 0 heads in 10 consecutive tosses? 5 heads out of 10? 6 out of 10? (3 pts)

(a) All three sequences are the same length, and every individual sequence of that length has the *same* probability. There are $2^{10} = 1024$ of them, so the probability of any one of them is $1/1024 = 0.0009765625$.

(b) There is $\binom{10}{0} = 1$ sequence with 0 heads, $\binom{10}{5} = 252$ sequences with 5 heads, and $\binom{10}{6} = 210$ sequences with 6 heads, so the probabilities are $1/1024$, $252/1024 = 0.2460938$ and $210/1024 = 0.2050781$. We can also do this using the binomial distribution, setting $n = 10$ and $p = 0.5$. (Notice that when $p = 0.5$, $p = 1 - p$, so $p^x(1-p)^{n-x} = p^n$.)

# 4 Dice

A fair die is one where all six faces are equally likely.
(a) Let $X$ = the number that comes up on a fair die. What is $\mathbf{E}[X]$? (5 pts)
(b) Let $X_1$, $X_2$ and $X_3$ be three independent fair die, and $S = X_1 + X_2 + X_3$. What is $\mathbf{E}[S]$? (2 pts)
(c) What is $\Pr(S = 18)$? (3 pts)

(a) $\mathbf{E}[X] = \sum_{i=1}^{6} i\frac{1}{6} = 3.5$

(b) The expectation of a sum is the sum of the expectations: $\mathbf{E}[S] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \mathbf{E}[X_3] = 3 \times 3.5 = 10.5$.

(c) The only way to get $S = 18$ is if $X_1 = X_2 = X_3 = 6$. So

$$\Pr(S = 18) = \Pr((X_1 = 6) \cap (X_2 = 6) \cap (X_3 = 6))$$

Since the dice are independent, the probabilities of intersections is just the product of the individual probabilities,

$$\Pr(S = 18) = \Pr(X_1 = 6)\Pr(X_2 = 6)\Pr(X_3 = 6) = \frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} = \frac{1}{216} = 0.00462963$$

## 5 A Continuous Random Variable

Consider a continuous random variable $X$ which lies between $-1$ and $+1$. The density of $X$ is proportional to $1 - x^2$.
(a) What is the density of $X$? (10 pts)
(b) What is the cumulative distribution function of $X$? (5 pts)
(c) Find the expectation of $X$. (5 pts) *Hint*: you do not need to integrate to do part (c).

(a) We are given that $f(x) \propto 1 - x^2$ when $-1 \leq x \leq 1$, and $= 0$ elsewhere. Since $f(x)$ has to integrate to 1 over all space, we know that

$$f(x) = \frac{1 - x^2}{\int_{-1}^{1} 1 - y^2 \, dy}$$

Finding the denominator is integration:

$$\int_{-1}^{1} 1 - y^2 \, dy = y - \left.\frac{y^3}{3}\right|_{-1}^{1} = \frac{4}{3}$$

so

$$f(x) = \frac{3}{4}\left(1 - x^2\right)$$

(b) Because $-1 \leq x \leq 1$, if $x < -1$ then $F(x) = 0$, and if $x > 1$ then $F(x) = 1$. In between, we integrate.

$$\begin{aligned} F(x) &= \int_{-1}^{x} f(y) \, dy \\ &= \frac{3}{4}\left[x - \frac{x^3}{3} + \frac{2}{3}\right] \end{aligned}$$

(c) Notice that $f(x)$ is symmetric about the origin, $f(x) = f(-x)$. So $\mathbf{E}[X] = \int_{-1}^{1} x f(x) \, dx$ will be zero — there will always be canceling contributions from $x$ and $-x$. We can do this by integration too:

$$\begin{aligned} \mathbf{E}[X] &= \int_{-1}^{1} x \frac{3}{4}(1 - x^2) \, dx \\ &= \frac{3}{4} \int_{-1}^{1} x - x^3 \, dx \\ &= \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{x=-1}^{x=1} \\ &= 0 \end{aligned}$$

## 6 Quality Control Charts

Your job is to monitor the concentration of xylene in the output of a municipal water-treatment plant. When the process is in control, the mean concentration of xylene is 6 parts per million, and the standard deviation per measurement is also 6 parts per million. Because measurements are small and cheap, you take $n = 36$ independent measurements every day.

(a) What is the standard error (standard deviation of the sample mean)? (5 pts.)

(b) What are the upper and lower control limits (UCL and LCL), at the usual "three sigma" level? (5 pts)

(c) The figure (next page) shows the daily sample mean concentrations of xylene over a month. Is the process in control? (5 pts)

(d) Is the distribution of the sample mean Gaussian? Explain. (5 pts)

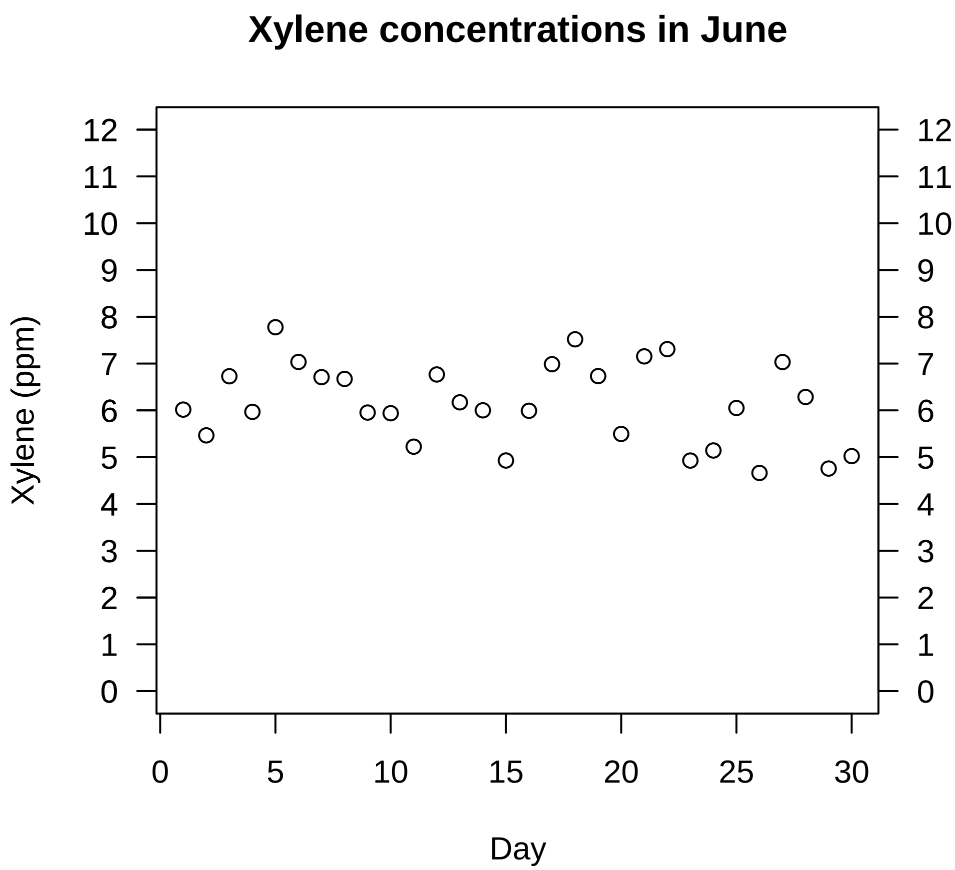
(a) The standard error is the population standard deviation $\sigma$ divided by $\sqrt{n}$, $n$ being the sample size; here $\sqrt{36} = 6$, and the s.e. is 1.

(b)

$$\begin{aligned} \text{UCL} &= \mu + 3\frac{\sigma}{\sqrt{n}} = 6 + 3 = 9 \\ \text{LCL} &= \mu - 3\frac{\sigma}{\sqrt{n}} = 6 - 3 = 3 \end{aligned}$$

(c) Drawing horizontal lines across the figure at 9 and 3 ppm, we see that none of the daily measurements are outside the lines; the process is in control.

(d) The problem does not say anything about the distribution of the individual concentration measurements, so we cannot assume they are Gaussian. However, it does say that measurements are independent, so the central limit theorem should apply if we take enough of them; generally "enough" is 30, and here we have 36, so it's reasonable to expect the sample means will in fact be Gaussian.




## 7 Quality Control and False Alarms

The output of an fully-functioning power plant has a Gaussian distribution, with a mean of 10 MW and a standard deviation of 0.5 MW. Upper and lower control limits are established as 11.0 MW and 9.0 MW respectively. If a certain component fails, the mean output drops to 9.0 MW, without changing the standard deviation.
(a) What is the probability that the output lies between the control limits, when the power plant is working properly? (10 pts)
(b) What is the probability that the output lies between the control limits during one of the partial failures? (10 pts)
(c) If the power plant works properly 99% of the time, what is the probability that going outside the control limits means a component has failed? (10 pts) *Hint 1*: Use Bayes's rule. *Hint 2*: What fraction of the time is the output outside the control limits?

(a) Notice that we are looking at individual measurements here, not samples. (Or, if you like, the sample size $n = 1$.)

$$
\begin{aligned}
\Pr(\text{within limits}|\text{in control}) &= \Pr(9.0 \le X \le 11.0) \\
&= \Pr\left(\frac{9.0 - 10.0}{0.5} \le Z \le \frac{11.0 - 10.0}{0.5}\right) \\
&= \Phi(2) - \Phi(-2) \\
&= 0.9544997
\end{aligned}
$$

(b) The limits are the same as in (a), but the mean has shifted.

$$
\begin{aligned}
\Pr(\text{within limits}|\text{not in control}) &= \Pr\left(\frac{9.0 - 9.0}{0.5} \le Z \le \frac{11.0 - 9.0}{0.5}\right) \\
&= \Phi(4) - \Phi(0) \\
&= 0.4999683
\end{aligned}
$$

(c) Let $F$ = the event that the power plant has failed, and $C$ = the event that the measurement is outside the control limits. We want $\Pr(F|C)$. In parts (a) and (b), we calculated $\Pr(C'|F)$ and $\Pr(C'|F')$. Use the first hint and set up Bayes's rule:

$$
\Pr(F|C) = \frac{\Pr(C|F)\Pr(F)}{\Pr(C)}
$$

$\Pr(F)$ is given as 0.01. $\Pr(C|F) = 1 - \Pr(C'|F) = 0.5000317$. That just leaves $\Pr(C)$, which is what the second hint tells us to calculate. Use the rule of total probability:

$$
\Pr(C) = \Pr(C|F)\Pr(F) + \Pr(C|F')\Pr(F')
$$

$\Pr(C|F') = 1 - \Pr(C'|F') = 0.0455003$. $\Pr(F') = 0.99$. So

$$
\begin{aligned}
\Pr(C) &= 0.05004561 \\
\Pr(F|C) &= 0.0999152
\end{aligned}
$$