MTH135/STA104: Probability

Homework # 7 Due: Tuesday, Nov 1, 2005

Prof. Robert Wolpert

1. For some number $c > 0$ the random variable X has a continuous probability distribution with density function

$$f(x) = c\,x, \qquad 0 < x < 4$$

(so $f(x) = 0$ for $x \notin (0,4)$); thus for any interval (a,b),

$$\mathsf{P}[a < X \le b] = \int_a^b f(x)\,dx$$

Please answer the following questions about X and its density function $f(x)$:

a) Find the value of $c > 0$: $c =$ __________________

$$1 = \mathsf{P}[0 < X \le 4] = \int_0^4 c\,x\,dx = c\,4^2/2 = 8c,$$

so evidently $c = \frac{1}{8}$.

b) Find the indicated probabilities:

- $\mathsf{P}[-2 < X \le 2] = \mathsf{P}[0 < X \le 2] = \frac{1}{8}2^2/2 = \frac{1}{4}$
- $\mathsf{P}[2 < X \le 5] = \mathsf{P}[2 < X \le 4] = \frac{1}{8}(4^2 - 2^2)/2 = \frac{3}{4}$
- $\mathsf{P}[|X - 2| > 1] = 1 - \mathsf{P}[1 < X \le 3] = 1 - \frac{1}{8}(3^2 - 1^2)/2 = \frac{1}{2}$

2. The life-times (in years) of cathode-ray tubes (CRTs) are random variables T that satisfy

$$\mathsf{P}[T > t] = \frac{100}{t^2}, \qquad t > 10$$

for every $t > 10$. Please answer the following questions:

a) The equation above doesn't mention $t < 10$, but it does include all the information needed to determine it. Find

$$\mathsf{P}[T \le 5] = 0, \text{ since } \mathsf{P}[T \le 10] = 1 - \mathsf{P}[T > 10] = 1 - 1 = 0$$

b) Find the pdf $f(t)$ for T; be careful to give it correctly at every $t \in \mathbb{R}$.

$$f(t) = \begin{cases} 200/t^3 & t > 10 \\ 0 & t \le 10 \end{cases}$$

c) Evaluate the expected value of T:

$$\mathsf{E}[T] = \int_{10}^{\infty} t\, 200/t^3\, dt = \int_{10}^{\infty} 200\, t^{-2}\, dt = 20$$

d) Find the conditional probability of failure in the next year for a working CRT that is now t years old for each $t \ge 10$:

$$\mathsf{P}[T \le t+1 \mid T > t] = 1 - \frac{\mathsf{P}[T > t+1]}{\mathsf{P}[T > t]} = 1 - \frac{100/(t+1)^2}{100/t^2} = 1 - \left(\frac{t}{t+1}\right)^2$$

Which is more likely to last another year, a 10-year old CRT or a 20-year old CRT?

The 20-year-old CRT has a better survival chance—

$$\frac{400}{441} \approx 90.7\% > 82.6\% = \frac{100}{121}$$

e) Find the (instantaneous) hazard,

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{200/t^3}{100/t^2} = 2/t, \qquad t > 10$$

Does it increase, decrease, or stay constant?

It decreases for $t > 10$

3. In our fishing example we found that the number of fish caught in t hours, X_t, had a Poisson probability distribution with mass function

$$\mathsf{P}[X_t = x] = \frac{(\lambda t)^x}{x!} e^{-\lambda t}, \qquad x = 0, 1, 2, \cdots$$

with mean $\mu = \lambda t$, if we catch λ fish per hour on average and if the numbers of fish caught in disjoint time intervals are independent.

Find the probability density function $f_k(t)$ for T_k, the time at which the k^{th} fish is caught. Hint: Express the *event* $T_k \le t$ in terms of X_t and use the Poisson probabilities above. Half-credit for the case $k = 1$ of the first fish caught.

$$f_k(t) = \underline{\hspace{4cm}}$$

First find the CDF:

$$\begin{aligned} F_1(t) &= \mathsf{P}[T_1 \le t] \\ &= \mathsf{P}[X_t \ge 1] \\ &= 1 - \mathsf{P}[X_t = 0] = 1 - e^{-\lambda t} \quad \text{for } t > 0, \quad = 0 \text{ for } t \le 0. \end{aligned}$$

Taking derivatives,

$$f_1(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0; \\ 0 & t \le 0. \end{cases}$$

4. If X is a positive random variable with pdf $f(x)$ and CDF $F(x)$, so

$$\begin{aligned} F(x) &= \mathsf{P}[X \le x] \\ &= 0, \qquad x < 0 \\ f(x) &= F'(x) \\ &= 0, \qquad x < 0 \end{aligned}$$

and if $Y \equiv \sqrt{X}$, find the CDF $G(y) = \Pr[Y \le y]$ and pdf $g(y) = G'(y)$ in terms of $f(x)$ and $F(x)$.

$$\begin{aligned} G(y) &= \mathsf{P}[\sqrt{X} \le y] \\ &= \mathsf{P}[X \le y^2], \qquad y \ge 0 \\ &= F(y^2), \quad y \ge 0; \\ g(y) &= 2y\, f(y^2), \quad y > 0 \end{aligned}$$

while $G(y) = g(y) = 0$ for $y < 0$.

5. For each question below, give an example or a brief explanation of why none is possible:

a) If X has a continuous distribution, can a function $Y = g(X)$ possibly have a discrete distribution with $\mathsf{P}[Y = y] > 0$ for some y?

Yes— for example, let X be uniformly distributed on $(0, 1)$ and let $g(x) = 1$ for $x \le \frac{1}{2}$, $g(x) = 0$ for $x > \frac{1}{2}$.

b) If X has a discrete distribution with finitely-many or countably-many possible values $\{x_j\}$, can a function $Y = g(X)$ possibly have a continuous distribution with density function $g(y)$, so $\mathsf{P}[a < Y \le b] = \int_a^b g(y)\, dy$ for all $a < b$?

No— if X has only countably-many values, then $Y = g(X)$ also has only countably-many values, and cannot have a density function.

c) Is it possible for the density function $f(x)$ for some random variable X to satisfy $f(x) > 1$ at any point x?

Sure. Let X be uniform on the interval $(0, \frac{1}{2})$, for examle; then its density function $f(x) = 2$ on $(0, \frac{1}{2})$ (and zero elsewhere).

d) Is it possible for a density function $f(x)$ to have a strictly positive lower bound on the positive half-line, *i.e.*, to satisfy $f(x) \ge \epsilon$ for every $x \in (0, \infty)$, for some fixed positive number $\epsilon > 0$?

No. This would imply that $\mathsf{P}[X < x] \ge x\epsilon$ for every $x > 0$, which would be bigger than one for $x > 1/\epsilon$, violating the rules about probabilities.

e) Is it possible for the cumulative distribution $F[x] = \mathsf{P}[X \le x]$ to be strictly increasing (*i.e.*, satisfy $F(a) < F(b)$ for every $a < b$) on the entire positive half-line $0 < x < \infty$?

Yes again— let X have density function e^{-x} for $x > 0$, for example, to find $F(x) = 1 - e^{-x}$ on $(0, \infty)$, strictly increasing.

6. Let X be uniformly distributed on the interval $[-1, 2]$ and let $Y = X^2$. Find the probability density function $f(y)$ for Y, correct at every point $y \in \mathbb{R}$.

Here $Y = g(x)$ with $f_x(x) = \frac{1}{3}, \quad -1 < x < 2$, $g(x) = x^2$, $g'(x) = 2x$, and $g^{-1}(y) = \pm\sqrt{y}$; thus

$$f(y) = \sum_{x:x^2=y} \frac{1}{|2x|}\frac{1}{3}\mathbf{1}_{(-1,2)}(x) = \begin{cases} 0 & y<0 \\ \frac{1}{3\sqrt{y}} & 0 \le y < 1 \\ \frac{1}{6\sqrt{y}} & 1 \le y < 4 \\ 0 & 4 \le y. \end{cases}$$

The answer may also be found by first computing the CDF,

$$F(y) = \mathsf{P}[X^2 \le y] = \begin{cases} 0 & y<0 \\ \frac{2\sqrt{y}}{3} & 0 \le y < 1 \\ \frac{\sqrt{y}}{3} & 1 \le y < 4 \\ 1 & 4 \le y, \end{cases}$$

from which the answer follows by differentiation.

7. Let X be a random variable with mean $\mathsf{E}[X] = \mu$ and variance $\mathsf{Var}[X] = \mathsf{E}[(X-\mu)^2] = \sigma^2$. Define a function $\phi(a)$ by

$$\phi(a) = \mathsf{E}[(X-a)^2]$$

(this is the "expected squared error" for our best guess a of what value X might take). Find the point a^* where $\phi(a)$ takes on its minimum value $\phi(a^*)$, and find what that minimum value is. HINT: It does *not* matter whether X is continuous or discrete— you should not need to do any integrals or sums.

Expanding,

$$\begin{aligned} \phi(a) &= \mathsf{E}[(X-a)^2] \\ &= \mathsf{E}[X^2 - 2Xa + a^2] \\ &= \mathsf{E}[X^2] - 2\,a\,\mathsf{E}[X] + a^2 \end{aligned}$$

Differentiating with respect to a,

$$\begin{aligned} \phi'(a) &= -2\mathsf{E}[X] + 2a \\ &= 0 \Leftrightarrow a = a^* \equiv \mathsf{E}[X], \end{aligned}$$

so $a^* = \mu$ and $\phi(a^*) = \mathsf{E}[(X-\mu)^2] = \sigma^2$.

8. Let X be a random variable with mean $\mathsf{E}[X] = \mu$ and variance $\mathsf{Var}[X] = \mathsf{E}[(X-\mu)^2] = \sigma^2$, with a continuous distribution with density function $f(x)$. Define a function $\psi(a)$ by

$$\psi(a) = \mathsf{E}|X - a|$$

(this is the "expected absolute error" for our best guess a of what value X might take). Find the point $a^\#$ where $\psi(a)$ takes on its minimum value $\psi(a^\#)$, and find what that minimum value is. Can you find an example where the minimizing value of $\psi(a)$ is not unique?

For any $a \in \mathbb{R}$,

$$\begin{aligned} \psi(a) &= \mathsf{E}|X - a| \\ &= \int_{x<a} (a-x)\, f(x)\, dx + \int_{a<x} (x-a)\, f(x)\, dx \end{aligned}$$

Differentiating with respect to a,

$$\begin{aligned} \psi'(a) &= \int_{x<a} f(x)\, dx - \int_{a<x} f(x)\, dx \\ &= \mathsf{P}[X < a] - \mathsf{P}[X > a] \\ &= 0 \Leftrightarrow \mathsf{P}[X < a] = \frac{1}{2} = \mathsf{P}[X > a] \end{aligned}$$

so $a^\# = m$ is a *median* of X. This is unique if $f(x) > 0$ for x near m, but not if $f(x) = 0$ on some interval with probability $\frac{1}{2}$ on each side— for example, not if X is drawn uniformly from the set $[-2,-1] \cup [1,2]$.