5. Continuous Random Variables

A r.v. X is <u>continuous</u> if it has a <u>density function</u> $f_X(x)$ satisfying

$$P(a \le X \le b) = \int_a^b f_X(t)\,dt \quad \text{for all } a, b$$

Density functions are nonnegative and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Important facts:

$$P(X = x) = \int_x^x f_X(t)\,dt = 0$$

$$P(x \le X \le x + \Delta x) = \int_x^{x+\Delta x} f(t)\,dt \approx f(x)\,\Delta x$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt$$

$$F_X'(x) = f_X(x)$$

Example 1a Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

(a) $\underline{1} = \int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{2} C(4x - 2x^2)\,dx$

$$= C\left(2x^2 - \frac{2}{3}x^3\right)\Big|_{x=0}^{x=2}$$

$$= C\left(\left(8 - \frac{16}{3}\right) - 0\right) = \frac{8}{3}C \Rightarrow C = \boxed{\frac{3}{8}}$$

(b) $P(X>1) = \int_{1}^{\infty} f(x)\,dx = \int_{1}^{2} C(4x - 2x^2)\,dx$

$$= \frac{3}{8}\left(2x^2 - \frac{2}{3}x^3\right)\Big|_{x=1}^{x=2}$$

$$= \frac{3}{8}\left(\left(8 - \frac{16}{3}\right) - \left(2 - \frac{2}{3}\right)\right)$$

$$= \frac{3}{8}\left(\frac{8}{3} - \frac{4}{3}\right) = \frac{3}{8}\cdot\frac{4}{3} = \boxed{\frac{1}{2}}$$

Example 1d If X is continuous with distribution function F_X and density function f_X, find the density function of $Y = 2X$.

$$F_Y(y) = P(Y \le y)$$

$$= P(2X \le y)$$

$$= P\left(X \le \frac{y}{2}\right)$$

$$= F_X\left(\frac{y}{2}\right)$$

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\left(F_X\left(\frac{y}{2}\right)\right)$$

$$= F_X'\left(\frac{y}{2}\right)\cdot\left(\frac{1}{2}\right)$$

$$= \boxed{\frac{1}{2} f_X\left(\frac{y}{2}\right)}$$

5.2 Expected Value

If X is continuous, then

$$E[X] = \int_{-\infty}^{\infty} t\, f_X(t)\, dt$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(t)\, f_X(t)\, dt$$

Useful formula:

If X is never negative, then

$$E[X] = \int_0^{\infty} P(X > t)\, dt$$

Special case:

$$\boxed{\text{If } R(X) = \{0, 1, 2, \dots\}, \text{ then} \quad E[X] = \sum_{n=1}^{\infty} P(X \geq n)}$$

Example 2a Find $E[X]$ when the density function of X is

$$f(x) = \begin{cases} 2x & \text{if } \quad 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x\,(2x)\, dx$$

$$= \int_0^1 2x^2\, dx = \frac{2}{3} x^3 \Big|_{x=0}^{x=1} = \boxed{\frac{2}{3}}$$

Example ~~2a~~ 2e Find ~~$E[X]$~~ $Var(X)$ when the density function of X is

$$f(x) = \begin{cases} 2x & \text{if } \quad 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

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

$E[X] = \frac{2}{3}$ by Expl 2a

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_0^1 x^2(2x)\,dx = \int_0^1 2x^3\,dx$$

$$= \frac{1}{2}x^4 \Big|_{x=0}^{x=1} = \frac{1}{2}$$

$$\mathrm{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2$$

$$= \frac{1}{2} - \frac{4}{9}$$

$$= \boxed{\frac{1}{18}}$$

5.3 Uniform Distribution

$X \sim \mathrm{Unif}(0,1)$ means X has density

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{o.w.} \end{cases}$$

$X \sim \mathrm{Unif}(\alpha, \beta)$ means X has density

$$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta, \\ 0 & \text{o.w.} \end{cases}$$

Expl 3a $X \sim \text{Unif}(\alpha, \beta)$. Find EX and $Var(X)$.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

$$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta, \\ 0 & \text{o.w.} \end{cases}$$

$$E[X] = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha}\, dx$$

$$= \frac{1}{\beta - \alpha} \cdot \frac{1}{2} x^2 \Big|_{x=\alpha}^{x=\beta} = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)}$$

$$= \frac{\cancel{(\beta - \alpha)}(\beta + \alpha)}{2\cancel{(\beta - \alpha)}}$$

$$= \boxed{\frac{\alpha + \beta}{2}}$$

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

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$

$$= \int_{\alpha}^{\beta} x^2 \frac{1}{\beta-\alpha}\,dx$$

$$= \frac{1}{\beta-\alpha} \cdot \frac{1}{3} x^3 \Big|_{x=\alpha}^{x=\beta}$$

$$= \frac{\beta^3-\alpha^3}{3(\beta-\alpha)} = \frac{\cancel{(\beta-\alpha)}(\beta^2+\alpha\beta+\alpha^2)}{3\cancel{(\beta-\alpha)}}$$

$$= \frac{\beta^2+\alpha\beta+\alpha^2}{3}$$

$$Var(X) = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \left(\frac{\beta+\alpha}{2}\right)^2$$

$$= \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{\beta^2+2\alpha\beta+\alpha^2}{4}$$

$$= \frac{4\beta^2+4\alpha\beta+4\alpha^2-3\beta^2-6\alpha\beta-3\alpha^2}{12}$$

$$= \frac{\beta^2-2\alpha\beta+\alpha^2}{12} = \boxed{\frac{(\beta-\alpha)^2}{12}}$$

Example 3b If X is uniformly distributed over $(0, 10)$, calculate the probability that (a) $X < 3$, (b) $X > 6$, and (c) $3 < X < 8$.

$$f_X(x) = \begin{cases} \frac{1}{10} & \text{if } 0 < x < 10, \\ 0 & \text{o.w.} \end{cases}$$

(a) $$P(X < 3) = \int_{-\infty}^{3} f_X(x)\,dx$$

$$= \int_0^3 \frac{1}{10}\,dx = \boxed{\frac{3}{10}}$$

(b) $$P(X > 6) = \int_6^{\infty} f_X(x)\,dx$$

$$= \int_6^{10} \frac{1}{10}\,dx = \frac{10-6}{10}$$

$$= \frac{4}{10} = \boxed{\frac{2}{5}}$$

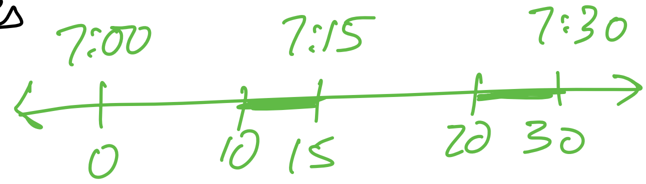
(c) $$P(3 < X < 8) = \int_3^8 f_X(x)\,dx = \int_3^8 \frac{1}{10}\,dx$$

$$= \frac{8-3}{10} = \frac{5}{10} = \boxed{\frac{1}{2}}$$

Example 3c Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

(a) less than 5 minutes for a bus;
(b) more than 10 minutes for a bus.

X = # of min. after 7:00 that the passenger arrives



$X \sim \text{Unif}(0, 30)$

(a) A = "He waits less than 5 minutes for a bus."

$$A = \{10 < X < 15\} \cup \{25 < X < 30\}$$

$$P(A) = P(10 < X < 15) + P(25 < X < 30)$$

$$= \int_{10}^{15} f_X(x)\,dx + \int_{25}^{30} f_X(x)\,dx$$

$$f_X(x) = \begin{cases} \frac{1}{30} & \text{if } 0 < x < 30, \\ 0 & \text{o.w.} \end{cases}$$

$$P(A) = \int_{10}^{15} \frac{1}{30}\,dx + \int_{25}^{30} \frac{1}{30}\,dx$$

$$= \frac{5}{30} + \frac{5}{30} = \frac{10}{30} = \boxed{\frac{1}{3}}$$

(b) B = "He waits more than 10 min."

$$B = \{0 < X < 5\} \cup \{15 < X < 20\}$$

$$P(B) = \int_0^5 \frac{1}{30}\,dx + \int_{15}^{20} \frac{1}{30}\,dx$$

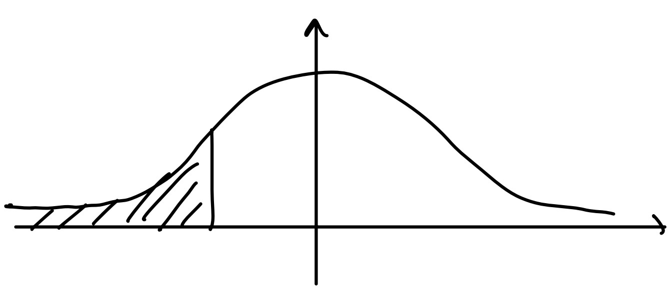
$$= \boxed{\frac{1}{3}}$$

5.4 Normal distribution

$X \sim \underbrace{N(0,1)}_{\text{standard normal distribution}}$ means X has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \leftarrow \text{integrates to } 1$$

$$F_X(x) = P(X \le x) = \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt}_{\Phi(x)} \quad \leftarrow \text{cannot calculate explicitly}$$



$$\Phi(0) = \frac{1}{2}$$

$$\Phi(-x) = 1 - \Phi(x)$$

Use table on p. 204 to calculate $\Phi(x)$

Often use Z for a standard normal r.v.

If $Z \sim N(0,1)$, then:

- $E[Z] = 0$
- $Var(Z) = 1$
- $P(|Z| \leq a) = P(-a \leq Z \leq a)$

$$= \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a))$$

$$= 2\Phi(a) - 1$$

$$P(|Z| \leq 1) = 2\Phi(1) - 1 \underset{\text{from table}}{\approx} 0.6826$$

$$P(|Z| \leq 2) \approx 0.9544$$

$$P(|Z| \leq 3) \approx 0.9974$$

$X \sim \underbrace{N(\mu, \sigma^2)}_{\text{(general) normal distribution}}$ means X has density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$E[X] = \mu$, $Var(X) = \sigma^2$

$$Z \sim N(0,1) \iff \mu + \sigma Z \sim N(\mu, \sigma^2)$$

$$X \sim N(\mu, \sigma^2) \iff \frac{X-\mu}{\sigma} \sim N(0,1)$$

If $X \sim N(\mu, \sigma^2)$, then

$$P(|X-\mu| \le \sigma) \approx 0.6826$$
$$P(|X-\mu| \le 2\sigma) \approx 0.9544$$
$$P(|X-\mu| \le 3\sigma) \approx 0.9974$$

Example 4b If X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$, find (a) $P\{2 < X < 5\}$; (b) $P\{X > 0\}$; (c) $P\{|X - 3| > 6\}$.

$$X \sim N(\underset{\mu}{3}, \underset{\sigma^2}{9})$$

$$X = \mu + \sigma \underset{N(0,1)}{Z}$$

$$X = 3 + 3Z$$

(a) $P(2 < X < 5) = P(2 < 3 + 3Z < 5)$

$= P(-1 < 3Z < 2)$

$= P\left(-\frac{1}{3} < Z < \frac{2}{3}\right)$

$$= \Phi\left(\tfrac{2}{3}\right) - \Phi\left(-\tfrac{1}{3}\right)$$

convert to decimals to use table

$$= \Phi\left(\tfrac{2}{3}\right) - \left(1 - \Phi\left(\tfrac{1}{3}\right)\right)$$

$$\approx \Phi(0.67) - (1 - \Phi(0.33))$$

from the table

$$\approx 0.7486 - (1 - 0.6293)$$

$$= 0.7486 - 0.3707$$

$$= \boxed{0.3779}$$

(b) $P(X > 0) = P(3 + 3Z > 0)$

$$= P(3Z > -3)$$

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

$$= 1 - P(Z \le -1)$$

$$= 1 - \Phi(-1)$$

$$= 1 - (1 - \Phi(1))$$

$$= \Phi(1) \approx \boxed{0.8413}$$

In general, $P(Z > -a) = \Phi(a)$ for exactly these reasons. You can use this property without showing these steps.

(c) $P(|X-3|>6)$

$= P(|3+3Z-3|>6)$

$= P(3|Z|>6)$

$= P(|Z|>2)$

$= P(Z>2) + P(Z<-2)$

$= 1-\Phi(2) + \Phi(-2)$

$= 1-\Phi(2) + (1-\Phi(2))$

$= 2(1-\Phi(2))$

$\approx 2(1-0.9772)$

$= 2(0.0228)$

$= \boxed{0.0456}$

Example 4d An expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with parameters $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?

- For concreteness, assume baby was born on Oct. 17 (290^{th} day of the year).
- The 50th day of the year is Feb. 19^{th}.

T = "Defendant is out of town Jan 1^{st} – Feb. 19^{th}."

F = "Defendant is the father."

G = "Baby was conceived before Jan 1^{st} or after Feb. 19^{th}."

X = Mother's gestation period (days)

$X \sim N(270, 100)$

$G = \{X > 290\} \cup \{X < 240\}$

They want: $P(G) = ?$

More interesting: $P(F|T) = ?$

Official question: $P(G) = ?$

$$X \sim \mathcal{N}(\underset{\mu}{270}, \underset{\sigma^2}{100})$$

$$X = \mu + \sigma \underbrace{Z}_{\sim N(0,1)}$$

$$X = 270 + 10Z$$

$$\begin{aligned}
P(G) &= P(X > 290) + P(X < 240) \\
&= P(270 + 10Z > 290) \\
&\quad + P(270 + 10Z < 240) \\
&= P(Z > 2) + P(Z < -3) \\
&= 1 - \Phi(2) + \cancel{\Phi(-3)} \quad 1 - \Phi(3) \\
&= 2 - \Phi(2) - \Phi(3) \\
&\approx 2 - 0.9772 - 0.9987 \\
&= 2 - 1.9759 \\
&= \boxed{0.0241}
\end{aligned}$$

Intuition check: If you were 98% sure he was the father before learning he was out of town, and $P(G) \approx 2.4\%$, how sure are you he's the father after learning he was out of town? What does your gut tell you?

But what does this have to do with $P(F|T)$?

$$O(F|T) = O(F)\cdot\frac{P(T|F)}{P(T|F^c)}$$

odds he's the father before learning he was out of town

odds he's the father *after* learning he was out of town.

Intuitively, we expect $O(F|T)$ to be a lot smaller that $O(F)$.

$$P(T\cap G|F) = P(T|F)\underbrace{\cancel{P(G|T\cap F)}}_{1}$$

(T and F implies G)

$$P(T\cap G|F) = P(G|F)\,P(T|F\cap G)$$

So $P(T|F) = P(G|F)\,P(T|F\cap G)$

Plug into top eqn

$$O(F|T) = O(F)\cdot\frac{P(T|F\cap G)}{P(T|F^c)}\cdot P(G|F)$$

If F^c, there's no reason he couldn't be out of town at that time.

If $F \cap G$, there's no reason he couldn't be out of town at that time.

So we might expect $P(T \mid F \cap G) \approx P(T \mid F^c)$

$$O(F \mid T) \approx O(F) \cdot P(G \mid F)$$

Knowing only F (and not T), the probability of G shouldn't change: $P(G \mid F) = P(G)$

$$O(F \mid T) \approx O(F) \cdot \underbrace{\cancel{P(G)}}_{0.0241}$$

Let's try some numbers to see what happens. Suppose we were 98% certain he's the father before learning he was out of town. So $P(F) = 0.98$. Then what is $P(F \mid T)$?

$$O(F) = \frac{P(F)}{P(F^c)} = \frac{0.98}{0.02} = 49$$

$$O(F|T) \approx 0.0241 \underset{49}{\cancel{O(F)}} = 1.1809$$

$$O(F|T) = \frac{P(F|T)}{P(F^c|T)}$$

$$1.1809 = \frac{P(F|T)}{1 - P(F|T)}$$

$$1.1809 - 1.1809\,P(F|T) = P(F|T)$$

$$P(F|T) = \frac{1.1809}{2.1809} \approx 54\%$$

After learning he was out of town, our certainty that he's the father drops from 98% to 54%.

Does this match your intuition from earlier?

Normal approx. to binomial

Let $0 < p < 1$. Let $S_n \sim \text{Binom}(n,p)$.

Let $Y_n = \dfrac{S_n - E[S_n]}{SD(S_n)} = \dfrac{S_n - np}{\sqrt{np(1-p)}}$.

Then

$$\lim_{n\to\infty} F_{Y_n}(x) = \Phi(x)$$

i.e. $P\left(\dfrac{S_n - np}{\sqrt{np(1-p)}} \le x\right) \xrightarrow{n\to\infty} \Phi(x)$

special case of central limit thm in Ch. 8

Intuitive version

If X is binomial w/ n large & p moderate, then

$$X \overset{d}{\approx} \underbrace{\mu}_{E[X]} + \underbrace{\sigma}_{SD(X)} \underbrace{Z}_{N(0,1)}$$

Often need a "continuity correction". Will see this in examples.

Example 4g Let X be the number of times that a fair coin that is flipped 40 times lands on heads. Find the probability that $X = 20$. Use the normal approximation and then compare it with the exact solution.

$X \sim \text{Binom}(40, \tfrac{1}{2})$

$P(X=20) = ?$

Normal approx:

$$\mu = np = 40\left(\tfrac{1}{2}\right) = 20$$

$$\sigma^2 = np(1-p) = 40\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = 10$$

$$X \overset{d}{\approx} \mu + \sigma \underbrace{Z}_{N(0,1)}$$

$$X \overset{d}{\approx} 20 + \sqrt{10}\, Z$$

First attempt:

$$P(X=20) \approx P(20 + \sqrt{10}\, Z = 20)$$

$$= P(Z=0) = 0 \quad (\text{b/c } Z \text{ is cont.})$$

Bad approximation!

X is discrete, so

$$X = 20 \iff 19.5 \le X \le 20.5$$

Use this!

continuity correction

$$P(X=20) \approx P(20 + \sqrt{10}\, Z \in [19.5, 20.5])$$

$$= P(19.5 \le 20 + \sqrt{10}\, Z \le 20.5)$$

$$= P(-0.5 \le \sqrt{10}\, Z \le 0.5)$$

$$= P\left(-\frac{1}{2\sqrt{10}} \le Z \le \frac{1}{2\sqrt{10}}\right)$$

$$= \Phi\left(\frac{1}{2\sqrt{10}}\right) - \cancel{\Phi\left(-\frac{1}{2\sqrt{10}}\right)}$$

$$1 - \Phi\left(\frac{1}{2\sqrt{10}}\right)$$

$$= 2\,\Phi\left(\frac{1}{2\sqrt{10}}\right) - 1$$

$$\approx 2\,\Phi(0.16) - 1$$

$$\approx 2(0.5636) - 1$$

$$= 1.1272 - 1$$

$$= \boxed{0.1272}$$

Exact:

$$P(X=20) = \binom{40}{20}\left(\frac{1}{2}\right)^{20}\left(\frac{1}{2}\right)^{40-20}$$

$$= \boxed{\frac{40!}{20!\,20!} \cdot \frac{1}{2^{40}}} = ?$$

Could use computing software to evaluate this to four decimal places. According to the textbook, it's 0.1254.

Example 4h The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.

$X =$ # of students that attend

$X \sim \text{Binom}(450, 0.3)$

$P(X > 150) \approx ?$

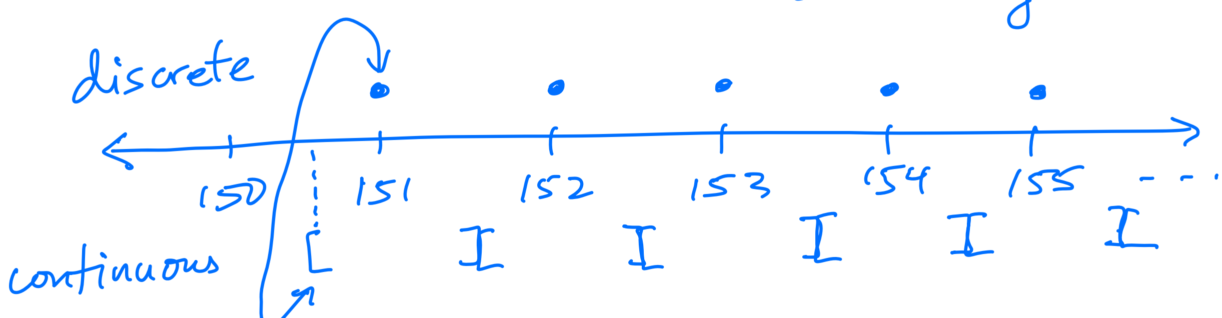
$\mu = np = 450(0.3) = 135$

$\sigma^2 = np(1-p) = 450(0.3)(0.7)$
$= 135(0.7) = 94.5$

$$X \overset{d}{\approx} \mu + \sigma \underset{N(0,1)}{Z} = 135 + \sqrt{94.5}\, Z$$

$$P(X > 150) \approx P(135 + \sqrt{94.5}\, Z > 150.5)$$

continuity correction



the point on the discrete side turns into the interval on the continuous side

$$P(X > 150) \approx P\left(Z > \frac{15.5}{\sqrt{94.5}}\right)$$

$$= 1 - \Phi\left(\frac{15.5}{\sqrt{94.5}}\right)$$

$$\approx 1 - \Phi(1.59)$$

$$\approx 1 - 0.9441$$

$$= \boxed{0.0559}$$

Example 4j Fifty-two percent of the residents of New York City are in favor of outlawing cigarette smoking on university campuses. Approximate the probability that more than 50 percent of a random sample of n people from New York are in favor of this prohibition when

(a) $n = 11$
(b) $n = 101$
(c) $n = 1001$

How large would n have to be to make this probability exceed .95?

X = # of people in sample that are in favor of outlawing cigarettes

$X \sim \text{Binom}(n, 0.52)$

$P\left(\frac{X}{n} > 0.5\right) \approx ?$

$\mu = np = 0.52n$

$\sigma^2 = np(1-p) = 0.52(0.48)n = 0.2496n$

$$\sigma = \sqrt{0.2496n}$$

$$X \overset{d}{\approx} \mu + \sigma Z = 0.52n + \sqrt{0.2496n}\, Z$$

Z ~ $N(0,1)$

n odd $\Rightarrow$ $n = 2k+1$

$$P\left(\frac{X}{n} > 0.5\right) = P(X > k)$$

cont. corr.

$$\approx P\left(0.52n + \sqrt{0.2496n}\, Z > k + 0.5\right)$$

$$= P\left(Z > \frac{k + 0.5 - 0.52n}{\sqrt{0.2496n}}\right)$$

$k = \frac{n-1}{2} = \frac{n}{2} - \frac{1}{2}$

$k + 0.5 = 0.5n$

$$= P\left(Z > -\frac{0.02n}{\sqrt{0.2496n}}\right)$$

$$= P\left(Z > -\frac{0.02}{\sqrt{0.2496}}\sqrt{n}\right)$$

Using the property that $P(Z > -a) = \Phi(a)$. See Expl 5.4b, Part (b).

$$= \Phi\left(\frac{0.02}{\sqrt{0.2496}}\sqrt{n}\right)$$

$$\approx \Phi(0.04\sqrt{n})$$

(a) $n = 11$: $\Phi(0.04\sqrt{11}) \approx \Phi(0.13) \approx \boxed{0.5517}$

(b) $n = 101$: $\Phi(0.04\sqrt{101}) \approx \Phi(0.40) \approx \boxed{0.6554}$

(c) $n = 1001$: $\Phi(0.04\sqrt{1001}) \approx \Phi(1.27) \approx \boxed{0.8980}$

Final question:

$$\Phi(0.04\sqrt{n}) \geq 0.95 \implies n = ?$$

$\Phi(1.64) \approx 0.9495$

$\Phi(1.65) \approx 0.9505$

best we can do with basic use of the table. Could be more accurate with a computer, or by using linear interpolation.

Need $0.04\sqrt{n} \geq 1.65$

$$n \geq \left(\frac{1.65}{0.04}\right)^2 = 1701.5625$$

So need $\boxed{n = 1702}$

5.5 Exponential dist.

$\lambda > 0$

$X \sim \mathrm{Exp}(\lambda)$ means X has density

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0, \\ 0 & \text{o.w.} \end{cases}$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x \lambda e^{-\lambda t}\,dt$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0, \\ 0 & \text{o.w.} \end{cases}$$

$$\boxed{P(X > t) = e^{-\lambda t} \quad \text{if } t > 0}$$

$$E[X] = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}$$

Time between events in a Poisson process has exponential dist.

Memoryless property:

$$P(X > s+t \mid X > t) = P(X > s)$$

Think of X as the "lifetime" of something
time 0 is its "birth", time X is its "death"

Example 5b Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = \frac{1}{10}$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait

(a) more than 10 minutes;
(b) between 10 and 20 minutes.

X = length of other person's call (min.)

$X \sim Exp\left(\frac{1}{10}\right)$

SOLN 1

(a) $P(X>10) = \int_{10}^{\infty} f_X(x)\,dx$

$$f_X(x) = \begin{cases} \frac{1}{10} e^{-\frac{1}{10}x} & \text{if } x \geq 0, \\ 0 & \text{o.w.} \end{cases}$$

$$P(X>10) = \int_{10}^{\infty} \frac{1}{10} e^{-\frac{1}{10}x}\,dx$$

$$= -e^{-\frac{1}{10}x} \Big|_{x=10}^{x=\infty} = 0-(-e^{-1}) = \boxed{e^{-1}}$$

(b) $P(10<X<20) = \int_{10}^{20} \frac{1}{10} e^{-\frac{1}{10}x}\,dx$

$$= -e^{-\frac{1}{10}x} \Big|_{x=10}^{x=20} = -e^{-2}-(-e^{-1}) = \boxed{e^{-1}-e^{-2}}$$

SOLN 2

Use $P(X>t) = e^{-\lambda t}$ when $X \sim Exp(\lambda)$

(a) $\lambda = \frac{1}{10}$, $P(X>10) = e^{-10\lambda} = \boxed{e^{-1}}$

(b) $P(10<X<20) = P(X>10) - P(X>20)$

$$= e^{-10\lambda} - e^{-20\lambda} = \boxed{e^{-1}-e^{-2}}$$

Example 5d Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5000-mile trip, what is the probability that he or she will be able to complete the trip without having to replace the car battery? What can be said when the distribution is not exponential?

X = lifetime of battery (miles)

$EX = 10000$

t = # of miles already on the battery

$P(\underline{X > 5000 + t} \mid \underline{X > t}) = ?$

we've already gone t miles and the battery's not dead yet.

what's the probability we can make it 5000 more miles?

(a) $X \sim Exp(\lambda)$

$$EX = \frac{1}{\lambda} = 10000 \Rightarrow \lambda = \frac{1}{10000}$$

$$P(X > 5000 + t \mid X > t) = P(X > 5000)$$

memoryless property

$$= e^{-5000\lambda} = \boxed{e^{-\frac{1}{2}}}$$

(b) X has dist. function F

The most we can say is

$$P(X > 5000 + t \mid X > t)$$

$$= \frac{P(\{X > 5000 + t\} \cap \{X > t\})}{P(X > t)}$$

$$= \frac{P(X > 5000 + t)}{P(X > t)}$$

$$= \boxed{\frac{1 - F(5000 + t)}{1 - F(t)}}$$

Need to know both F and t to get a numerical answer.

Hazard rates

Let X be a r.v. that's always positive.

Think of it as the "lifetime" of something.

X is not necessarily exponential.

X has density f and dist. func. F.

Define

$$\boxed{\lambda(t) = \frac{f(t)}{1-F(t)}}$$

hazard rate of X

Meaning:

$$P(X\in(t,t+\Delta t)\mid X>t) = \frac{P(X\in(t,t+\Delta t))}{P(X>t)}$$

$$= \frac{1}{1-F(t)}\int_t^{t+\Delta t} f(s)\,ds$$

$$\approx \frac{1}{1-F(t)} f(t)\,\Delta t = \lambda(t)\,\Delta t$$

Can recover dist. func. from hazard rate:

$$F(t) = \begin{cases} 1-e^{-\int_0^t \lambda(s)\,ds} & \text{if } t>0, \\ 0 & \text{o.w.} \end{cases}$$

OR

$$\boxed{P(X>t) = e^{-\int_0^t \lambda(s)\,ds} \quad \text{if } t>0}$$

If $\lambda(t)=\lambda$ is a const. func., then $X\sim Exp(\lambda)$.

Example 5f One often hears that the death rate of a person who smokes is, at each age, twice that of a nonsmoker. What does this mean? Does it mean that a nonsmoker has twice the probability of surviving a given number of years as does a smoker of the same age?

X = lifetime of a particular smoker (years)

Y = lifetime of a particular nonsmoker (years)

$$\lambda_X(t) = 2\lambda_Y(t)$$

$P(X > s+t \mid X > t)$ = probability that the smoker of age t survives an additional s years

$P(Y > s+t \mid Y > t)$ = probability that the nonsmoker of age t survives an additional s years

$$P(Y > s+t \mid Y > t) \overset{?}{=} 2P(X > s+t \mid X > t)$$

comma is commonly used for intersection

$$P(Y > s+t \mid Y > t) = \frac{P(Y > s+t,\ Y > t)}{P(Y > t)}$$

$$= \frac{P(Y > s+t)}{P(Y > t)}$$

$$P(Y>t) = \exp\left(-\int_0^t \lambda_Y(u)\,du\right)$$

$$P(Y>s+t) = \exp\left(-\int_0^{s+t} \lambda_Y(u)\,du\right)$$

$$= \exp\left(-\int_0^t \lambda_Y(u)\,du - \int_t^{s+t} \lambda_Y(u)\,du\right)$$

$$= \underbrace{\cancel{\exp\left(-\int_0^t \lambda_Y(u)\,du\right)}}_{P(Y>t)} \exp\left(-\int_t^{s+t} \lambda_Y(u)\,du\right)$$

So...

$$P(Y>s+t \mid Y>t) = \exp\left(-\int_t^{s+t} \lambda_Y(u)\,du\right)$$

and

$$P(X>s+t \mid X>t) = \exp\left(-\int_t^{s+t} \lambda_X(u)\,du\right)$$

using the given

$$= \exp\left(-\int_t^{s+t} 2\lambda_Y(u)\,du\right)$$

$$= \left(\exp\left(-\int_t^{s+t} \lambda_Y(u)\,du\right)\right)^2$$

$$= \left(P(Y>s+t \mid Y>t)\right)^2$$

$$P(Y > s+t \mid Y > t) = \sqrt{P(X > s+t \mid X > t)}$$

So the nonsmoker's chance of survival is not double the smoker's. It's the square root of the smoker's.

E.g. if the smoker had a 50% chance of survival, the nonsmoker has a $\sqrt{\frac{1}{2}} \approx 0.707 = 70.7\%$ chance of survival.

If the smoker's death rate were 3x the nonsmoker's, we'd use cube root, and so on.

HW: Ch.5: 1, 3, 4, 7, 8, 11, 13, 17, 21, 22, 25, 26, 29, 31, 32, 37, 39

36, 38