Stat 343, Thu 1-Oct-2020 -- Thu 1-Oct-2020

Probability and Statistics

Fall 2020

Thursday, October 1st 2020

Wk 5, Th

Topic:: Exponential distributions

Read:: FASt 3.1

Continuous Distributions

Example: A first pdf. Let

$$f(x) \;=\; \begin{cases} ax(1-x), & x \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$

(a) Draw a graph of f in RStudio, using $a = 2$ for convenience.

(b) Determine the value of a so that

$$\int_{-\infty}^{\infty} f(x)\,dx \;=\; 1.$$

Definition 1: A function $f\colon \mathbb{R} \to \mathbb{R}$ is a **probability density function**, or **pdf**, if it has the properties

- $f(x) \geq 0$ for all $x \in \mathbb{R}$, and
- $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$

Example: Is the following function a pdf for some choice of a?

$$f(x) \;=\; \begin{cases} 0, & x < 0 \\ ae^{-2x}, & x \geq 0 \end{cases}$$

What if we replace (-2) with $b \leq 0$?

A continuous random variable X

- can take on values throughout an interval
- satisfies $P(X = x) = 0$
- has a **cumulative distribution function**, or **cdf**, defined to be $F_X(x) \;=\; P(X \leq x)$. Note that F is
  - monotone increasing (in x), making F (almost everywhere) differentiable (a deep insight from 20th Century analysis).
  - the derivative $f = F'$ is (almost everywhere) nonnegative, and

$$F(x) \;=\; P(X = x) \;=\; \int_{-\infty}^{x} f(t)\,dt.$$

  - as a consequence of the above, $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx \;=\; \lim_{x\to\infty} F(x) = 1.$

So, the derivative of F is a pdf.

- satisfies $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)\,dx$.

As a result, one can *define* a random variable X

- starting with a pdf and using integration to get its cdf, or
- starting with a cdf (any $F\colon \mathbb{R} \to [0,1]$ with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$), and using differentiation to get its pdf.

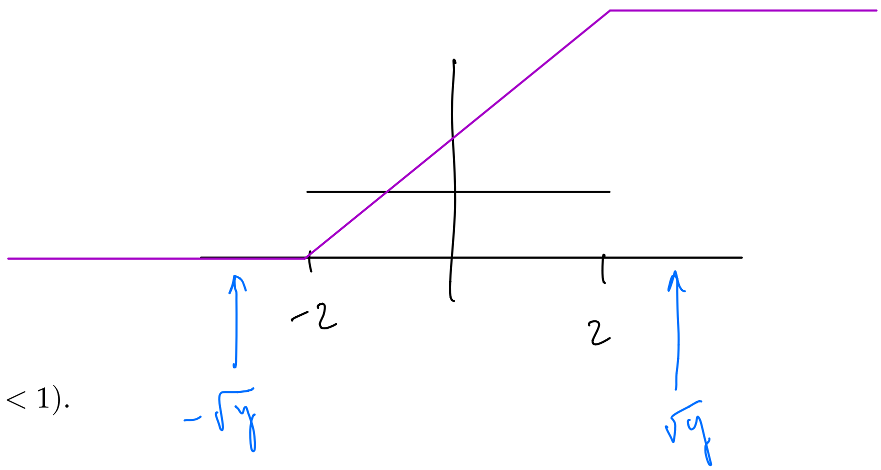
Uniform distributions

This family of distributions arises from having a pdf that is an appropriately-scaled indicator function on a finite interval $[a,b]$. That is, $X \sim \text{Unif}(a,b)$ if

$$f(x) \;=\; \left\{ \begin{array}{ll} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{array} \right\} \;=\; \frac{1}{b-a}\,\chi_{[a,b]}(x) \;=\; \frac{1}{b-a}\,[\![a \leq x \leq b]\!].$$

Example $X \sim \text{Unif}[-2,2]$.

(a) Plot the pdf $f_X(x)$.



(b) Find $P(X < -3)$, $P(X \leq 0)$, and $P(X < 1)$.

(c) Use commands in R to redo part (b).

(d) Give a formula for the cdf $F_X(x)$.

$$F_X(x) = \begin{cases} 0, & \text{if } x < -2 \\ \frac{1}{4}(x+2), & -2 \leq x \leq 2 \\ 1, & \text{if } x > 2 \end{cases}$$

(e) Suppose $Y = 5X$ Find the cdf $F_Y(y)$ and pdf $f_Y(y)$.

cdf for Y, $F_Y(y) = P(Y \leq y) = P(5X \leq y)$

$= P(X \leq \frac{1}{5}y) = F_X(\frac{1}{5}y)$

$$= \begin{cases} 0, & \text{if } y < -10 \\ \frac{1}{4}(\frac{1}{5}y + 2) & \text{if } -10 \leq y \leq 10 \\ 1, & \text{if } y > 10 \end{cases}$$

pdf for Y: $$f_Y(y) = \begin{cases} 1/20, & -10 < y < 10 \\ 0, & \text{otherwise} \end{cases}$$

(f) Suppose $Y = X^2$ Find the cdf $F_Y(y)$ and pdf $f_Y(y)$.

cdf for Y $F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$

Note: $P(X^2 \leq y) = 0$ if $y < 0$.

For $y \geq 0$, $P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

$$= \begin{cases} 0, & \text{if } y < 0 \\ \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{4} dx, & 0 < y < 4 \\ 1, & \text{if } y \geq 4 \end{cases} = \begin{cases} 0, & \text{if } y \leq 0 \\ \frac{\sqrt{y}}{2}, & 0 < y < 4 \\ 1, & y \geq 4 \end{cases}$$

Exponential distributions

Following an example above, we have that, for $\lambda > 0$, the function

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

is a pdf. Its corresponding cdf is given by

$$F(x) = \int_{-\infty}^{x} f(t;\lambda)\,dt = \left\{ \begin{array}{ll} \int_0^x \lambda e^{-\lambda t}\,dt, & x \geq 0 \\ 0, & \text{otherwise} \end{array} \right\} = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$x \to \infty$

A continuous random variable X for which $P(X \leq x) = F(x)$ as above is said to have an exponential distribution with **rate** parameter λ, and we write $X \sim \text{Exp}(\lambda)$.

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Note that if $X \sim \text{Exp}(\lambda)$ and $b > a > 0$, then

$$\begin{aligned} P(X > b \mid X > a) &= \frac{P(X > a, X > b)}{P(X > a)} = \frac{P(X > b)}{P(X > a)} = \frac{1 - (1 - e^{-\lambda b})}{1 - (1 - e^{-\lambda a})} \\ &= \frac{e^{-\lambda b}}{e^{-\lambda a}} = e^{-\lambda(b-a)} = P(X > b - a), \end{aligned}$$

which can be interpreted as the same **memoryless** phenomenon as observed in a geometric random variable.

Example: Life of lightbulbs. Suppose $X \sim \text{Exp}(1/1000)$ is a random variable that models the working lifetime (in hours) of a certain lightbulb.

(a) Describe a use for the R command

```
rexp(6, rate=1/1000)

[1] 1164.11739 3376.43934   14.96474  945.27346 2682.52012 2104.05491
```

You might try out this set of commands:

```
simLifetimes <- rexp(5000, rate=1/1000)
gf_dhistogram(~ simLifetimes) %>%
  gf_dist("exp", params = list(rate = 1/1000))
mean(~ simLifetimes)
```

Can you discover/guess why `gf_dhistogram()` instead of `gf_histogram()`?

(b) Estimate the probability of one of these lightbulbs lasting more than 2000 hours.

1 - pexp(2000, 1/1000)

Task: Propose a formula for the expected value of a continuous r.v. X.

- Past experience involves expected values for discrete r.v.s X with pmf $f_X(x)$: $\text{E}(X) = \sum_x x f_X(x)$. What is an appropriate analog for a continuous r.v. X with pdf $f_X(x)$?

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

cont. r.v.

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du$$

How would we define $\text{Var}(X)$ when X is a continuous r.v.?

$$\text{Var}(X) = E\left((X-\mu_X)^2\right)$$

Question: Does the formula $\text{Var}(X) = \text{E}(X^2) - [\text{E}(X)]^2$ still hold?

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^{\infty} x\lambda e^{-\lambda x}\,dx$$

- Compute $\text{E}(X)$ when $X \sim \text{Exp}(\lambda)$.

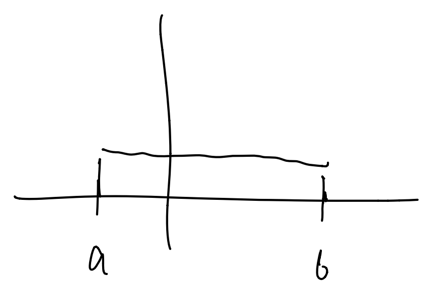
$$= \lambda \int_0^{\infty} x e^{-\lambda x}\,dx$$

$u = x$, $du = dx$

$dv = e^{-\lambda x}dx$, $v = -\frac{1}{\lambda}e^{-\lambda x}$

$$= \frac{1}{\lambda}.$$

- Compute $\text{E}(X)$ when $X \sim \text{Unif}(a, b)$.



- Compute $\text{Var}(X)$ when $X \sim \text{Unif}(a, b)$.

$X \sim \text{Unif}(a,b)$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b = \frac{1}{b-a}\cdot\frac{1}{2}(b^2-a^2) = \frac{(b+a)\cancel{(b-a)}}{2\cancel{(b-a)}} = \frac{1}{2}(a+b).$$

$$E(X^2) =$$

Exponential and Poisson distributions

Suppose $X \sim \text{Pois}(\lambda)$, a Poisson random variable. Recall that X models the counting of random events in a unit of time, that λ is the expected value (average number) in that time interval, and the pmf is

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

Now let Y be the time until the next occurrence, a continuous random variable. If $F_Y(y)$ represents the cdf for Y, then if y is measured in the same time units as λ, then λy is the average number of events that occur in a time interval $[0, y]$, and

$$\begin{aligned} F_Y(y) &= P(Y \le y) = 1 - P(Y > y) = 1 - P(X = 0 \text{ in time interval } [0, y]) \\ &= 1 - e^{-\lambda y}\frac{\lambda^0}{0!} = 1 - e^{-\lambda y}, \end{aligned}$$

$\leftarrow$ matches cdf found yesterday(?) for exponential r.v.

showing that $Y \sim \text{Exp}(\lambda)$.

$P(Y > y)$ gives probability that gap between events exceeds y

$= P(\text{no events witnessed in } [0, y])$

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

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - (1/\lambda)^2 = \frac{1}{\lambda^2}.$$

```
-------------------------------
Thursday, October 01st 2020
-------------------------------
Due::    PS05 due at 6 pm
```

Say we're watching arrivals at queue at bank.

rate parameter $\lambda = 30$ (30 arrivals per hour on avg.)

$X \sim \text{pois}(\lambda = 30)$

$P(X = 20 \text{ in 1 hour}) = \text{dpois}(20, \text{lambda} = 30)$

$P(X \le 10 \text{ in 20 minutes}) = \text{ppois}(10, 10)$

$P(X = 0 \text{ in 10 mins.}) = \text{dpois}(0, 5) = 0.00\_$

If Y measures gap between arrivals

$P(Y > 10 \text{ mins.}) = 1 - \text{pexp}(1, 5) = 0.43$

rate parameter where 1 unit of time is 10 mins

So, rate parameter appropriate for 1 min.

$$\lambda = 1/2$$

$$P(Y > 10 \text{ mins}) = 1 - \text{pexp}(10, 1/2)$$