STAT 712 fa 2022 Lec 6 slides

Suite of ought-to-know probability distributions

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

1. Discrete distributions

2. Continuous distributions

3. Exponential families

Distributions related to Bernoulli trials

Let $X_i = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$ $\quad i = 1, 2, \dots$ indicate outcomes of indep. Bernoulli trials.

Jakob Bernoulli

1. $Y = X_1 =$ success of single trial $\implies Y \sim \text{Bernoulli}(p)$,

$$p_Y(y) = p^y(1-p)^{1-y} \cdot \mathbf{1}(y \in \{0,1\})$$

2. $Y = \sum_{i=1}^n X_i =$ # successes in n trials $\implies Y \sim \text{Binomial}(n, p)$,

$$p_Y(y) = \binom{n}{y} p^y (1-p)^{n-y} \cdot \mathbf{1}(y \in \{0, 1, 2, \dots, n\})$$

3. $Y = \min\{i : X_i = 1\} =$ # of trial of 1st success $\implies Y \sim \text{Geometric(p)}$,

$\underbrace{FF\dots F}_{y-1}\,S$ ← y

$$p_Y(y) = (1-p)^{y-1} p \cdot \mathbf{1}(y \in \{1, 2, \dots\})$$

4. $Y = \min\{i : \sum_{j=1}^i X_j = r\} =$ # of trial of rth succ. $\implies Y \sim \text{negBin}(r, p)$,

$\underbrace{\boxed{FFSFSSFF\dots}}_{y-1,\ r-1 \text{ successes}}\,S$ ← y^{th}

$$p_Y(y) = \binom{y-1}{r-1}(1-p)^{y-r} p^r \cdot \mathbf{1}(y \in \{r, r+1, r+2, \dots\})$$

r

Exercise: Derive these pmfs; then find mean, variance, and mgf of each.

Hypergeometric distribution

Draw $K \geq 0$ marbles from a bag of $N \geq 0$ marbles, of which $M \geq 0$ are red.

Then $X = \#$ red marbles drawn $\implies X \sim \text{Hypergeometric}(N, M, K)$, has pmf

$$p_X(x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \cdot \mathbf{1}(x \in \{\max(K-(N-M), 0), \dots, \min(K, M)\}).$$

$0, \dots, K$

$\mathbb{E}X = K\frac{M}{N}$

Exercise: For $X \sim \text{Hypergeometric}(N, M, K)$, find $\mathbb{E}X$ and $\operatorname{Var} X$.

Assum $K \leq M$, $K \leq N-M$ so $\mathcal{X} = \{0, \dots, K\}$

$$\mathbb{E}X = \sum_{x=0}^{K} x \, \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

$$= \sum_{x=1}^{K} x \, \frac{\dfrac{M!}{x!\,(M-x)!} \, \dfrac{(N-M)!}{(K-x)!\,((N-M)-(K-x))!}}{\dfrac{N!}{K!\,(N-K)!}}$$

$$= K\frac{M}{N} \underbrace{\sum_{x=1}^{K} \frac{\dfrac{(M-1)!}{(x-1)!\,(M-x)!} \, \dfrac{(N-M)!}{(K-x)!\,((N-M)-(K-x))!}}{\dfrac{(N-1)!}{(K-1)!\,(N-K)!}}}_{=1\ ?}$$

$$\binom{(N-1)-(M-1)}{(K-1)-(x-1)}$$

$$\overbrace{\binom{M-1}{x-1}} \qquad \overbrace{\frac{\left((N-1)-(M-1)\right)!}{\left((K-1)-(x-1)\right)!\,\left((N-1)-(M-1)-\left((K-1)-(x-1)\right)\right)!}}$$

$$= K\frac{M}{N}\sum_{x=1}^{K}\underbrace{\frac{\dfrac{(M-1)!}{(x-1)!\,\left((M-1)-(x-1)\right)!}\;\dfrac{(N-M)!}{(K-x)!\,\left((N-M)-(K-x)\right)!}}{\dfrac{(N-1)!}{(K-1)!\,(N-K)!}}}$$

$$(N-K)! = \left((N-1)-(K-1)\right)! \qquad \binom{N-1}{K-1}$$

$$= K\frac{M}{N}\underbrace{\left(\cancel{\sum_{x=1}^{K}}\;\sum_{y=0}^{K-1}\frac{\binom{M-1}{y}\binom{(N-1)-(M-1)}{(K-1)-y}}{\binom{N-1}{K-1}}\right)}_{\overset{?}{=}1}$$

$=1$

pmf of Hypergeom$(N-1, M-1, K-1)$

$$= K\frac{M}{N}.$$

Discrete uniform distribution and empirical distribution

1. If X takes the values $1, \dots, K$ with prob. $1/K$ then $X \sim \text{discUnif}(K)$.

$$p_X(x) = \frac{1}{K} \cdot \mathbf{1}(x \in \{1, \dots, K\})$$

2. If X takes the values $\overbrace{x_1, \dots, x_n}^{\text{data}}$ with prob. $1/n$ then $X \sim \text{empDist}(x_1, \dots, x_n)$.

$$p_X(x) = \frac{1}{n} \cdot \mathbf{1}(x \in \{x_1, \dots, x_n\})$$

Exercise: Give the mean and variance of $X \sim \text{empDist}(x_1, \dots, x_n)$.

$$\mathbb{E}X = \sum_{x \in \{x_1, \dots, x_n\}} x \frac{1}{n} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \bar{x}_n$$

$$\mathrm{Var}\, X = \mathbb{E}\left(X - \mathbb{E}X\right)^2$$

$$= \sum_{x \in \{x_1, \dots, x_n\}} \left(x - \bar{x}_n\right)^2 \cdot \frac{1}{n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \bar{x}_n\right)^2$$

Poisson distribution

For some $\lambda > 0$, let $Y_n \sim \text{Binomial}(n, \underbrace{\lambda/n}_{p})$ for $n = 1, 2, \dots$ Then

$$\boxed{\lim_{n \to \infty} M_{Y_n}(t)} = M_Y(t) = e^{\lambda(e^t - 1)} \quad \text{for all } t \in \mathbb{R},$$

where $M_Y(t)$ is the mgf of $Y \sim \text{Poisson}(\lambda)$, which has pmf

$$p_Y(y) = \frac{e^{-\lambda}\lambda^y}{y!} \cdot \mathbf{1}(y \in \{0, 1, 2, \dots\}).$$

So for large n, $Y_n \sim \text{Binomial}(n, \lambda/n)$ behaves like $Y \sim \text{Poisson}(\lambda)$.

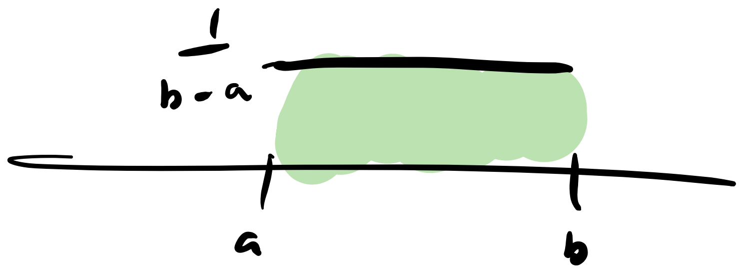
Often posited for # of occurrences of an event per unit of time/space, where

- the events occur independently from one another, and
- are as likely to occur in any time/space interval as in any other.

$\mathbb{E}X = \lambda$ $\quad$ $\text{Var}\, X = \lambda$

Exercise: Find mean and variance and derive mgf.

1. Discrete distributions
2. Continuous distributions
3. Exponential families

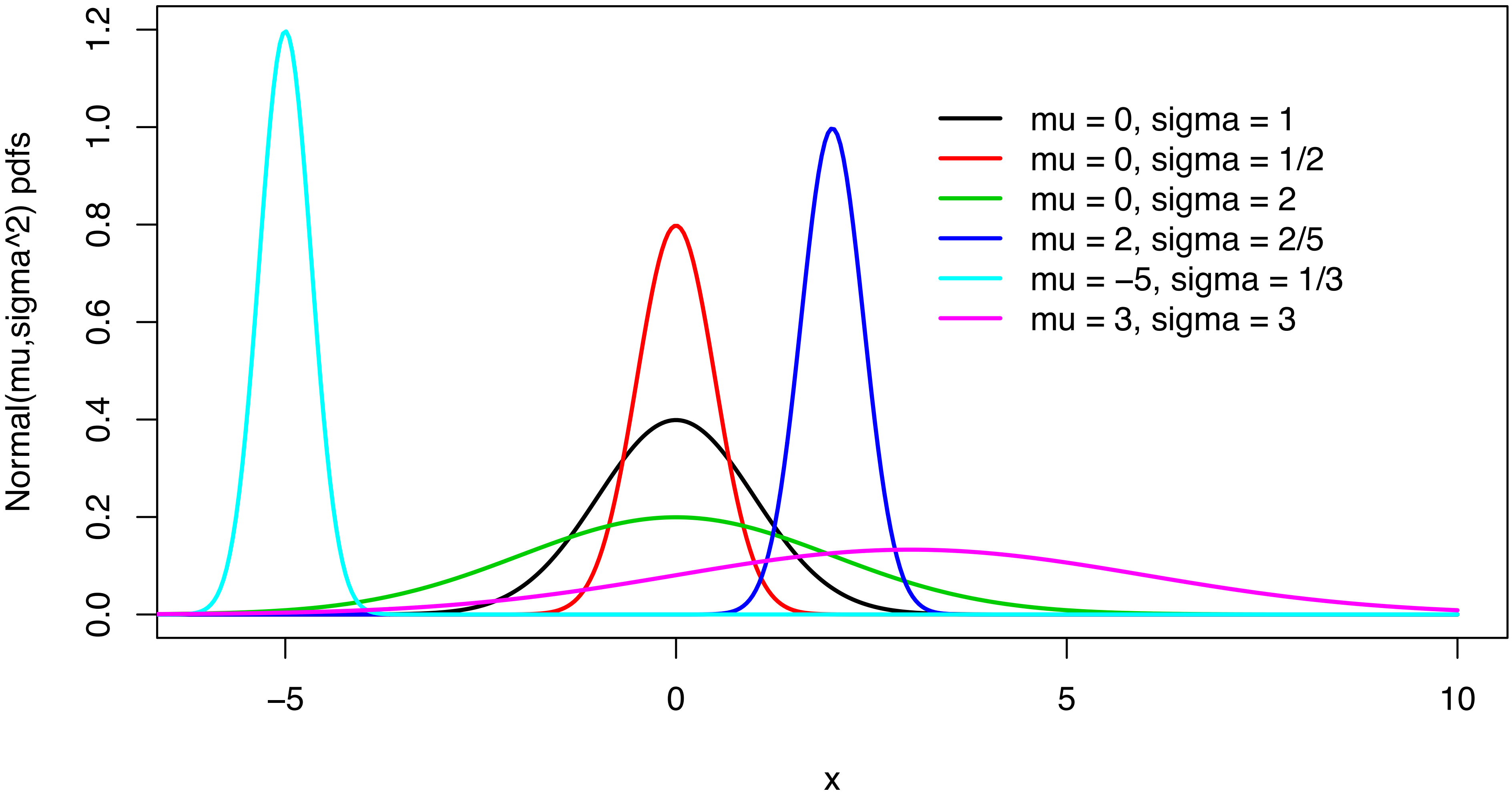


- The pdf of the $\boxed{\text{Uniform}(a, b)}$ distribution is given by

$$f_X(x; a, b) = \frac{1}{b-a}\mathbf{1}(a < x < b) \quad \text{for } x \in \mathbb{R},$$

  for $a < b$.
- Parameters:
  - a is the lower bound of the support
  - b is the upper bound of the support
- The Uniform$(0, 1)$ pdf is $f_X(x) = \mathbf{1}(0 < x < 1)$.

pdfs of several Normal distributions



- The pdf of the Normal(μ, σ^2) distribution is given by

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{for } x \in \mathbb{R}.$$

- Parameters:
  - $\mu \in \mathbb{R}$ is a *location parameter*
  - $\sigma > 0$ is a *scale parameter*
- If $X \sim$ Normal(μ, σ^2), then
  - $\mathbb{E}X = \mu$
  - $\text{Var}\, X = \sigma^2$
  - X has mgf $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$ for all $t \in \mathbb{R}$

"standard Normal Distribution"

- The pdf and cdf of the Normal$(0,1)$ distribution get special notation:

$$\boxed{\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}}$$

$\phi(t)$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt \quad \text{for } z \in \mathbb{R}.$$

$Z \overset{d}{=} \frac{X-\mu}{\sigma}$ standard Normal

Exercises: Let $\boxed{X \sim \text{Normal}(\mu, \sigma^2)}$ and $Z \sim \text{Normal}(0, 1)$. Show

1. $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$ for all $t \in \mathbb{R}$
2. $X \overset{d}{=} \sigma Z + \mu$
3. $\boxed{\mathbb{E}Z = 0}$ ✓
4. $\text{Var}\, Z = 1$
5. $\mathbb{E}X = \mu$
6. $\text{Var}\, X = \sigma^2$
7. $\int_{-\infty}^{\infty} \phi(z) dz = 1$

(2) We have $M_Z(t) = e^{t^2/2}$

$$M_{\sigma Z + \mu}(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2} = e^{\mu t + \sigma^2 t^2/2}.$$

(3) $\mathbb{E}Z = \int_{\mathbb{R}} z \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = M_Z^{(1)}(0)$

$$M_Z^{(1)}(0) = \frac{d}{dt} e^{t^2/2}\Big|_{t=0} = e^{t^2/2} t \Big|_{t=0} = 0$$

① $M_X(t) = \mathbb{E}\, e^{tX}$

$$= \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} + tx \right] dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{x^2 - 2x\mu + \mu^2 - tx2\sigma^2}{2\sigma^2} \right] dx$$

"complete the square"

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{\left\{x^2 - 2x(\mu + t\sigma^2) + (\mu + t\sigma^2)^2\right\} - (\mu + t\sigma^2)^2 + \mu^2}{2\sigma^2} \right] dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{\left(x - (\mu + t\sigma^2)\right)^2 \overbrace{- (\mu + t\sigma^2)^2 + \mu^2}^{\cancel{\mu^2} - \cancel{\mu^2} - 2\mu t\sigma^2 - t^2\sigma^4}}{2\sigma^2} \right] dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{\left(x - (\mu + t\sigma^2)\right)^2}{2\sigma^2} + \frac{2\mu t\sigma^2 + t^2\sigma^4}{2\sigma^2} \right] dx$$

$$= e^{\mu t + \sigma^2 t^2/2} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{\left(x - (\mu + t\sigma^2)\right)^2}{2\sigma^2} \right] dx}_{=1 \quad \text{pdf of } N(\mu + t\sigma^2, \sigma^2)}$$

$$= e^{\mu t + \sigma^2 t^2/2}$$

(4) $\mathrm{Var}\, Z = \mathbb{E} Z^2 - (\underbrace{\mathbb{E} Z}_{=0})^2 = \mathbb{E} Z^2$

$$M_Z^{(2)}(0) = \frac{d^2}{dt^2} e^{t^2/2}\Big|_{t=0} = \frac{d}{dt} e^{t^2/2} t\Big|_{t=0} = e^{t^2/2}\cdot 1 + e^{t^2/2} t\cdot t\Big|_{t=0}$$

$$= 1.$$

$\mathrm{Var}\, Z = 1.$

Let $X = \sigma Z + \mu \sim N(\mu, \sigma^2).$

(5) $\mathbb{E} X = \mathbb{E}(\sigma Z + \mu) = \sigma \underbrace{\mathbb{E} Z}_{=0} + \mu = \mu$

(6) $\mathrm{Var}\, X = \mathrm{Var}(\sigma Z + \mu) = \sigma^2 \underbrace{\mathrm{Var}\, Z}_{=1} = \sigma^2$

(7) $\int_{\mathbb{R}} \phi(z)\, dz \overset{?}{=} 1$

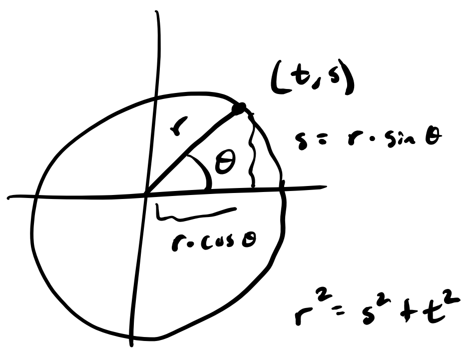
$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = 1$$

$$\Longleftrightarrow \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz\right)^2 = 1$$

$$\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz\right)^2 = \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt\right)\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-s^2/2}\, ds\right)$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{2\pi} e^{-\frac{t^2+s^2}{2}}\, ds\, dt$$

? look in calculus book.



$r^2 = s^2 + t^2$

$\vdots$

$$= \int_0^{2\pi}\int_0^{\infty} \frac{1}{2\pi} e^{-\frac{r^2}{2}}\, r\, dr\, d\theta$$

$$= \int_0^{2\pi} \frac{1}{2\pi} \underbrace{\int_0^{\infty} r e^{-r^2/2}\, dr}\, d\theta$$

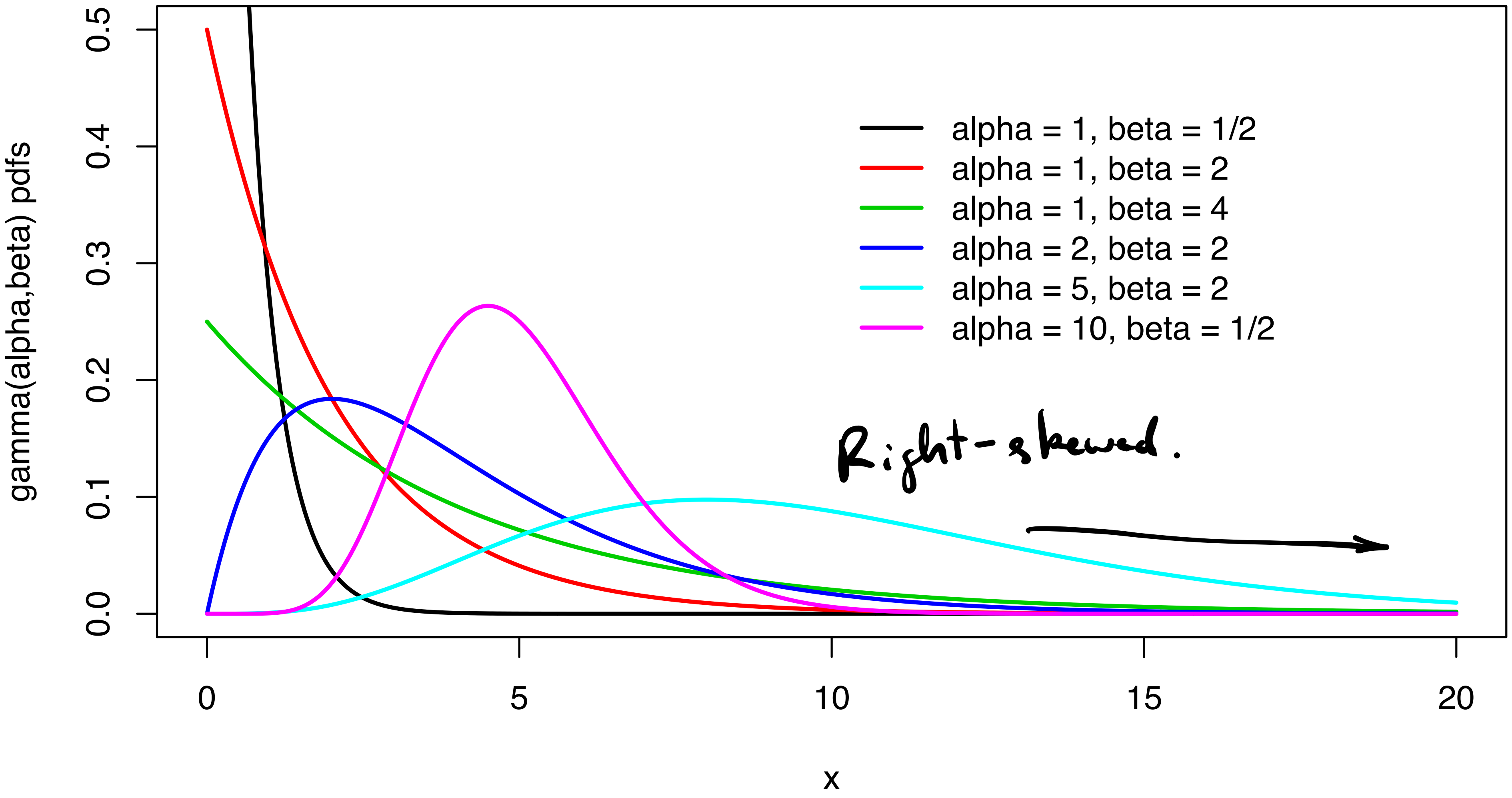
$$\frac{d}{dr} - e^{-r^2/2} = r e^{-r^2/2}$$

$$= \int_0^{2\pi} \frac{1}{2\pi} \underbrace{\left[-e^{-r^2/2}\right]\Big|_0^{\infty}}_{(0-(-1))}\, d\theta$$

$$= \int_0^{2\pi} \frac{1}{2\pi}\, d\theta$$

$$= 1$$

pdfs of several Gamma distributions



- The pdf of the Gamma(α, β) distribution is given by

$$f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left[-\frac{x}{\beta}\right] \quad \text{for } x > 0.$$

$\Gamma(\alpha)$

Gamma function.

- Parameters:
  - $\alpha > 0$ is a *shape parameter*
  - $\beta > 0$ is a *scale parameter*
- If $X \sim$ Gamma(α, β), then
  - $\mathbb{E}X = \alpha\beta$
  - $\text{Var}\, X = \alpha\beta^2$
  - X has mgf $M_X(t) = (1 - \beta t)^{-\alpha}$ for $t < 1/\beta$.

$\alpha = 1$

$\Gamma(1) = 1$

$$F_X(x) = \begin{cases} \int_0^x \frac{1}{\Gamma(\alpha)\beta^\alpha} t^{\alpha-1} e^{-t/\beta}\, dt & = \text{pgamma}(x, \text{shape} = \alpha, \text{scale} = \beta) \\ 0 & x \leq 0 \end{cases}$$

The gamma distributions are brought to you by the gamma function.

Gamma function

For any $\alpha \in \mathbb{C}$ with $\text{Re}(\alpha) > 0$, the *gamma function* is given by

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du.$$

always real for us.

These are some of its properties:

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for all $\text{Re}(\alpha) > 0$.
2. $\Gamma(n) = (n-1)!$ for any integer $n > 0$.
3. $\Gamma(1/2) = \sqrt{\pi}$.

Exercise:

1. Prove the above properties.
2. Show $\mathbb{E}X = \alpha\beta$ if $X \sim \text{Gamma}(\alpha, \beta)$.

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u}\, du$$

Integration by parts

$$\int u\, dv = uv - \int v\, du$$

① $$\Gamma(\alpha+1) = \int_0^\infty \underbrace{x^{\overbrace{(\alpha+1)-1}^{=\alpha}}}_{u} \underbrace{e^{-x}\, dx}_{dv}$$

$$\frac{du}{dx} = \alpha x^{\alpha-1}, \quad du = \alpha x^{\alpha-1} dx$$

$$v = -e^{-x}$$

$$= \underbrace{x^{(\alpha+1)-1}(-e^{-x})\Big|_0^\infty}_{=0} - \int_0^\infty (-e^{-x})\, \alpha x^{\alpha-1} dx$$

$$= \alpha \underbrace{\int_0^\infty x^{\alpha-1} e^{-x}\, dx}_{\Gamma(\alpha)}$$

$$= \alpha\, \Gamma(\alpha)$$

②

$$\Gamma(1) = \int_0^\infty u^{1-1} e^{-u}\, du = \int_0^\infty e^{-u}\, du = -e^{-u}\Big|_0^\infty = (0-(-1)) = 1$$

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1$$

$$\Gamma(3) = 2\,\Gamma(2) = 2 \cdot 1$$

$$\Gamma(4) = 3\,\Gamma(3) = 3 \cdot 2 \cdot 1 \quad \ldots$$

$$\Gamma(n) = (n-1)!$$

(3) $\Gamma(1/2) = \int_0^\infty u^{1/2-1} e^{-u} \, du$

$= \int_0^\infty u^{-1/2} e^{-u} \, du$ $\quad u = \frac{z^2}{2}, \; du = z\,dz$

$z = \sqrt{2u}$

$= \int_0^\infty \left(\frac{z^2}{2}\right)^{-1/2} e^{-z^2/2} \, z \, dz$

$= \sqrt{2} \int_0^\infty e^{-z^2/2} \, dz$

$= \sqrt{2} \; \frac{1}{2} \int_{-\infty}^{\infty} e^{-z^2/2} \, dz$

pdf $N(0,1)$

$= \sqrt{2} \; \frac{1}{2} \; \sqrt{2\pi} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz}_{=1}$

$= \sqrt{\pi}$

Suppose $X \sim \text{Gamma}(\alpha, \beta = 1)$,

then $X \sim f_X(x) = c \cdot x^{\alpha-1} e^{-x} \, I(x>0)$

$$1 = \int_0^\infty c \cdot x^{\alpha-1} e^{-x} \, dx$$

$$= c \underbrace{\int_0^\infty x^{\alpha-1} e^{-x} \, dx}_{\Gamma(\alpha)} \iff c = \frac{1}{\Gamma(\alpha)}$$

$$X \sim \text{Gamma}(\alpha, \beta)$$

$$X \sim f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \mathbb{1}(x>0)$$

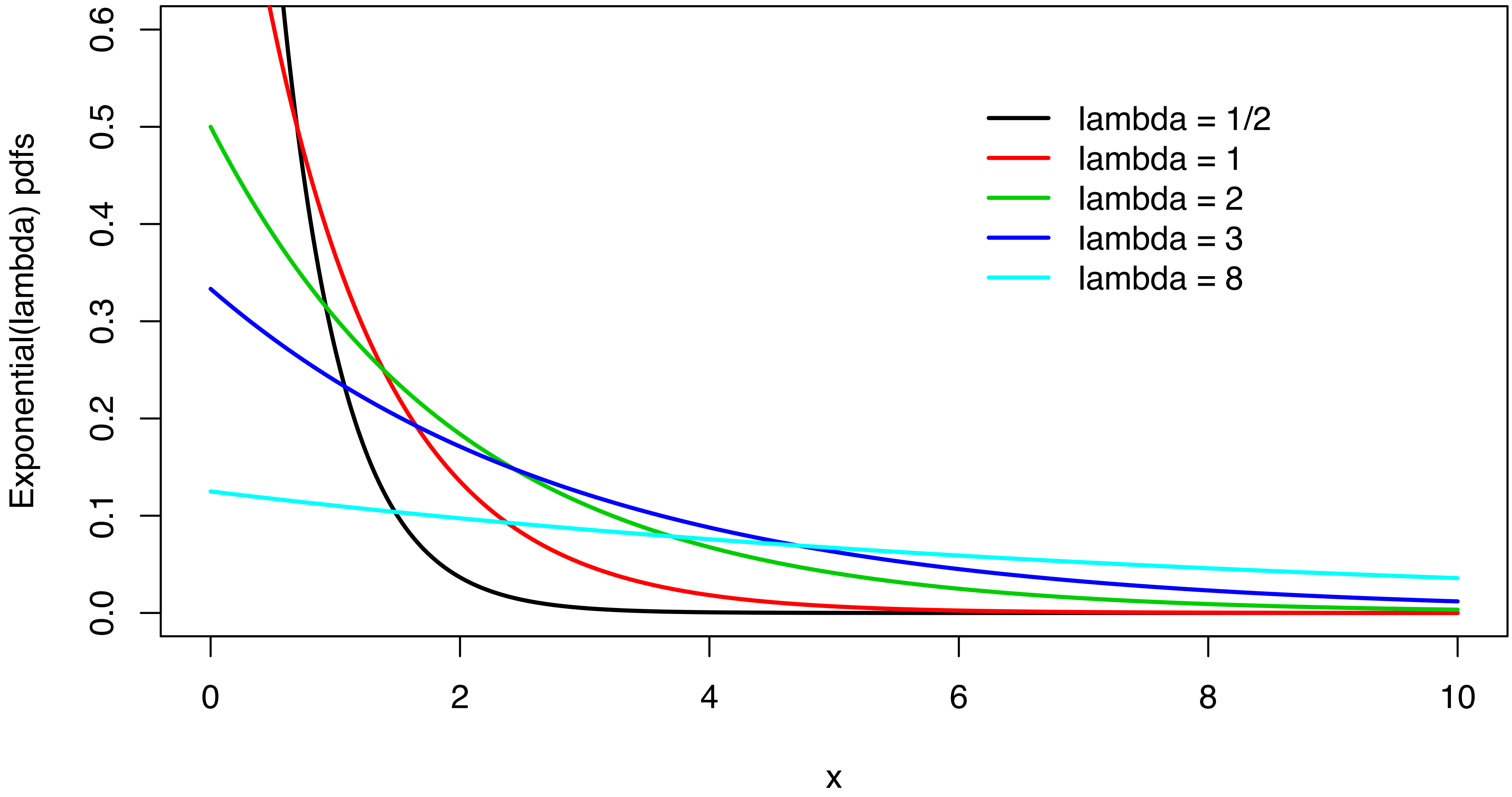
$$\mathbb{E}X = \int_0^\infty x \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \, dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha+1)\beta^{\alpha+1} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^{(\alpha+1)-1} e^{-x/\beta} \, dx}_{=1}$$

$$= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha}$$

$$= \frac{\alpha \cancel{\Gamma(\alpha)} \beta}{\cancel{\Gamma(\alpha)}} = \alpha\beta.$$

pdfs of several Exponential distributions



- The pdf of the Exponential(λ) distribution is given by

$$f_X(x;\lambda) = \frac{1}{\lambda}\exp\left[-\frac{x}{\lambda}\right] \quad \text{for } x > 0.$$

- Parameter:
  - $\lambda > 0$ is a *scale parameter*
- If $X \sim$ Exponential(λ), then
  - $\mathbb{E}X = \lambda$
  - $\operatorname{Var} X = \lambda^2$
  - X has mgf $M_X(t) = (1-\lambda t)^{-1}$ for $t < 1/\lambda$.
  - $X \sim$ Gamma$(1, \lambda)$

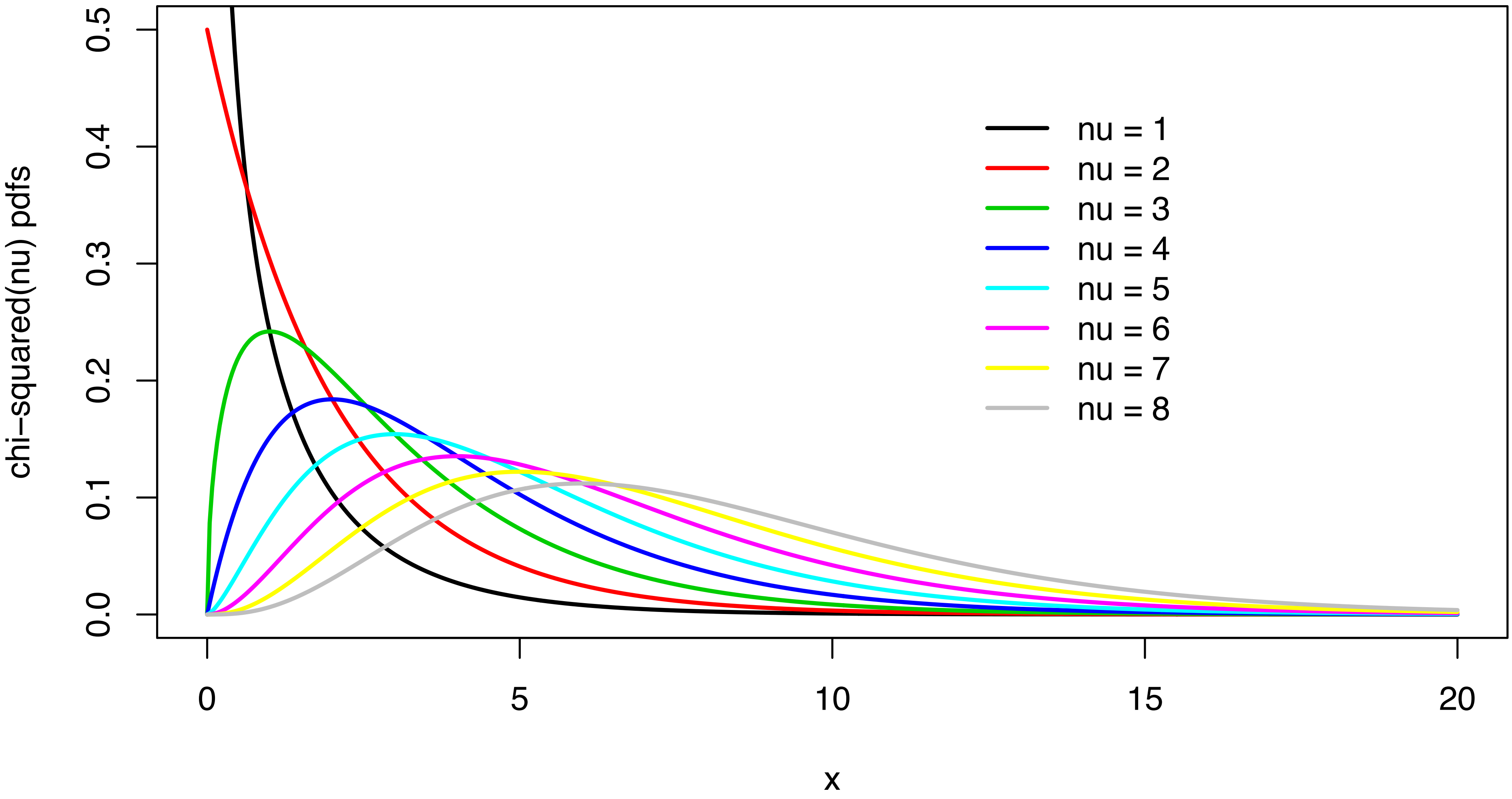
Exercise: Find the $\boxed{\text{cdf}}$ of the Exponential(λ) distribution.

$X \sim f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} \mathbb{1}(x>0)$

$$F_X(x) = \begin{cases} \int_0^x \frac{1}{\lambda} e^{-t/\lambda}\, dt & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$= \begin{cases} 1 - e^{-x/\lambda}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

pdfs of several Chi-squared distributions



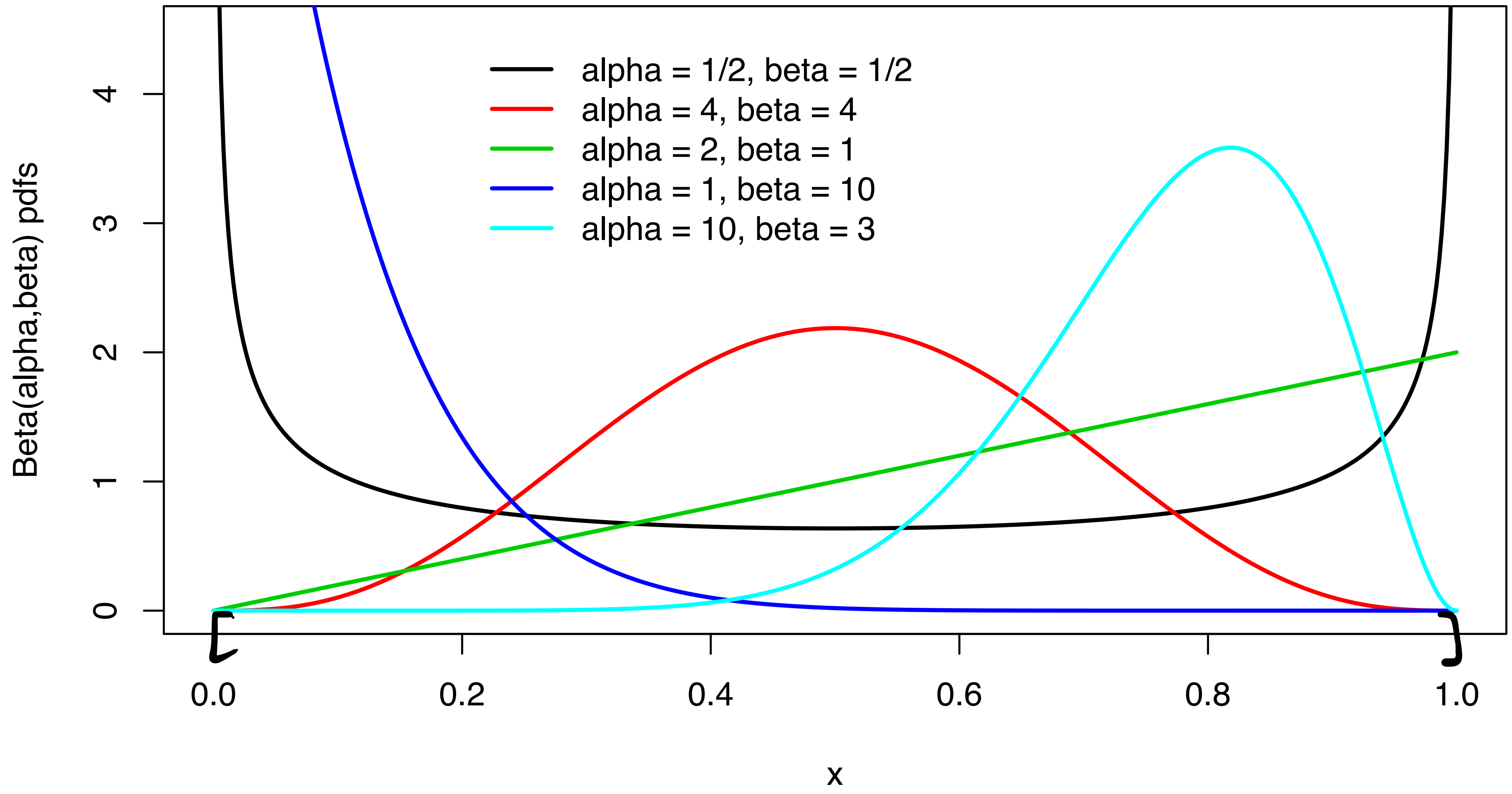
$\nu =$ "nu"

- The pdf of the Chi-squared(ν) distribution is given by

$$f_X(x;\nu) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} \exp\left[-\frac{x}{2}\right] \qquad \text{for } x > 0.$$

- Parameter:
  - $\nu > 0$ is called the *degrees of freedom*
- If $X \sim$ Chi-squared(ν), then
  - $\mathbb{E}X = \nu$
  - $\text{Var}\, X = 2\nu$
  - X has mgf $M_X(t) = (1-2t)^{-\nu/2}$ for $t < 1/2$.
  - $X \sim \text{Gamma}(\underbrace{\nu/2, 2}_{\alpha = \nu/2,\ \beta = 2})$

pdfs of several Beta distributions



- The pdf of the Beta(α, β) distribution is given by

$$f_X(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } x \in (0,1).$$

- Parameter:
  - $\alpha > 0$ is a *shape parameter*
  - $\beta > 0$ is a *shape parameter*
- If $X \sim \text{Beta}(\alpha, \beta)$, then
  - $\mathbb{E}X = \dfrac{\alpha}{\alpha+\beta}$
  - $\text{Var}\, X = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
  - X has mgf $M_X(t) = 1 + \sum_{k=1}^{\infty} \frac{t^k}{k!}\left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r}\right)$ for all $t \in \mathbb{R}$.

$\mathbb{E}\, e^{tX}$

- The Beta$(1,1)$ distribution is the Uniform$(0,1)$ distribution.

The beta distributions are brought to you by the beta function.

Beta function

For any $\alpha, \beta \in \mathbb{C}$ with $\text{Re}(\alpha) > 0$ and $\text{Re}(\beta) > 0$, the *beta function* is given by

$$B(\alpha, \beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Exercise: Show $\mathbb{E}X = \alpha/(\alpha+\beta)$ if $X \sim \text{Beta}(\alpha, \beta)$.

$$\mathbb{E}X = \int_0^2 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx$$

$$= \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)}} \underbrace{\int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\,\Gamma(\beta)} x^{(\alpha+1)-1} (1-x)^{\beta-1}}_{\text{int. over pdf of Beta}(\alpha+1,\,\beta),\ =1} dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\alpha\,\Gamma(\alpha)}{(\alpha+\beta)\,\Gamma(\alpha+\beta)}$$

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

We will now start talking about *families of distributions*.

Parametric family

For a pdf/pmf $f(\cdot;\theta)$ with some parameters $\theta = (\theta_1, \dots, \theta_d)^T \in \Theta \subset \mathbb{R}^d$, $d \geq 1$, we call the set of pdfs/pmfs

$$\{f(\cdot;\theta) : \theta \in \Theta\}$$

a *parametric family* of pdfs/pmfs.

Example: The Beta parametric family is the set of pdfs

$$\left\{ f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}(0 < x < 1) : \alpha > 0, \beta > 0 \right\}.$$

Exponential family

A parametric family $\{f(\cdot;\theta) : \theta \in \Theta\}$ is called an *exponential family* if each member can be written as

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right), \quad x \in \mathbb{R}$$

for some real-valued functions

- $h(\cdot) \geq 0$ and $t_1(\cdot), \ldots, t_k(\cdot)$ not depending on θ and
- $c(\cdot) \geq 0$ and $w_1(\cdot), \ldots, w_k(\cdot)$ not depending on x.

Why though??

- Many common families of distributions are exponential families.
- Generalized linear models are based on exponential family distributions.

Exercise: Show that the Normal(μ, σ^2) pdfs for $\mu \in \mathbb{R}$ and $\sigma > 0$ are an exponential family.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{\left[-\frac{x^2 - 2x\mu + \mu^2}{2\sigma^2}\right]}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[\left(-\frac{1}{2\sigma^2}\right)x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right]$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \underbrace{\frac{1}{\sigma} e^{-\mu^2/2\sigma^2}}_{c(\mu,\sigma^2)} \exp\left[\underbrace{\left(-\frac{1}{\sigma^2}\right)}_{w_1(\mu,\sigma^2) = -\frac{1}{\sigma^2}} \underbrace{\frac{x^2}{2}}_{t_1(x) = \frac{x^2}{2}} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2}} \underbrace{x}_{t_2(x) = x}\right]$$

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right),$$

$k=2$

$k=2$

$$t_1(x)w_1(\mu,\sigma^2) + t_2(x)w_2(\mu,\sigma^2)$$

Yes, it's an exponential family.

$k=1$

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right),$$

Exercise: Show that the Poisson(λ) pmfs for $\lambda > 0$ are an exponential family.

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \mathbb{1}(x \in \{0,1,2,\dots\})$$

$$= \underbrace{e^{-\lambda}}_{c(\lambda)} \underbrace{\frac{1}{x!}\mathbb{1}(x \in \{0,1,2,\dots\})}_{h(x)} \exp\Big(\underbrace{x}_{t_1(x)} \cdot \underbrace{\log\lambda}_{w_1(\lambda)}\Big).$$

Curved versus full exponential families

An exponential family $\{f(\cdot;\theta) : \theta \in \Theta\}$ is a *curved exponential family* if $d = \dim(\Theta) < k$, with k from the representation

d parameter

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right), \quad x \in \mathbb{R}.$$

If $d = k$, then $\{f(\cdot;\theta) : \theta \in \Theta\}$ is a *full exponential family*.

Examples:

1. The Beta$(\alpha, 2\alpha)$ pdfs for all $\alpha > 0$.
2. The Normal(μ, μ^2) pdfs for all $\mu \in \mathbb{R}$.

$N(\mu, \mu^2)$ "Curved exp. family"

$$= \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \underbrace{\frac{1}{\sigma} e^{-\frac{1}{2}}}_{c(\mu)} \exp\left[\underbrace{\left(-\frac{1}{\mu^2}\right)}_{w_1(\mu) = -\frac{1}{\mu^2}} \underbrace{\frac{x^2}{2}}_{t_1(x) = \frac{x^2}{2}} + \underbrace{\frac{1}{\mu}}_{w_2(\mu) = \frac{1}{\mu}} \underbrace{x}_{t_2(x) = x} \right]$$

$t_2(x) = x$

$t_1(x) = \frac{x^2}{2}$

$w_2(\mu) = \frac{1}{\mu}$

$w_1(\mu) = -\frac{1}{\mu^2}$