Continuous Distributions

The Gamma dist.

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

pdf $f(x) = \dfrac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$ for $x > 0$.

$(\alpha > 0, \beta > 0)$

$EX = \alpha\beta$

$\text{Var}X = \alpha\beta^2$

mgf $M_X(t) = \left(\dfrac{1}{1-\beta t}\right)^\alpha$, $t < \frac{1}{\beta}$.

Calculation of moments done earlier.

Derivation of mgf:

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

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

$$= \frac{1}{\beta^\alpha} \int_0^\infty \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x\left(\frac{1}{\beta}-t\right)}\,dx$$

which is finite when $t < \frac{1}{\beta}$

Now let $\frac{1}{\beta'} = \frac{1}{\beta} - t$

so that $\beta' = \dfrac{1}{\frac{1}{\beta} - t} = \dfrac{\beta}{1-\beta t}$.

Then integral becomes

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

$$= \left(\frac{\beta'}{\beta}\right)^\alpha \cdot 1 = \left(\frac{1}{1-\beta t}\right)^\alpha \text{ for } t < \frac{1}{\beta}.$$

<u>Closure Property</u>

If X_1, X_2 independent

$X_1 \sim \text{Gamma}(\alpha_1, \beta)$,

$X_2 \sim \text{Gamma}(\alpha_2, \beta)$,

then $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$

<u>Proof</u>: $M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t)$

$$= \left(\frac{1}{1-\beta t}\right)^{\alpha_1} \left(\frac{1}{1-\beta t}\right)^{\alpha_2} = \left(\frac{1}{1-\beta t}\right)^{\alpha_1+\alpha_2}$$

= mgf of $\text{Gamma}(\alpha_1 + \alpha_2, \beta)$. QED.

The Exponential Distribution ($\text{Exp}(\beta)$)

pdf $f_X(x) = \frac{1}{\beta} e^{-x/\beta}$ for $x > 0$ ($\beta > 0$),

$EX = \beta$, $\text{Var}\, X = \beta^2$,

cdf $F_X(x) = 1 - e^{-x/\beta}$ for $x > 0$,

$$P(X > x) = e^{-x/\beta} \text{ for } x > 0,$$

mgf $M_X(t) = \frac{1}{1-\beta t}$ for $t < 1/\beta$,

$\text{Exp}(\beta)$ same as $\text{Gamma}(1, \beta)$.

The (Continuous) Memoryless Property

If $X \sim \text{Exp}(\beta)$, then

$$P(X - t > s \mid X > t) = P(X > s) \qquad (*)$$

for all $s, t > 0$

or equivalently,

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

for all $s, t > 0$

Converse : If the r.v. Y satisfies $(*)$, then $Y \sim \text{Exp}(\beta)$ for some value of β.

Failure rate (hazard rate)

Suppose X has pdf $f(x)$, cdf $F(x)$,
and $f(x) = 0$ for $x < 0$.

Define the hazard function

$$h(t) = \lim_{\delta \downarrow 0} \frac{P(t < X \leq t + \delta \mid X > t)}{\delta}$$

$$\left(= \frac{f(t)}{1 - F(t)} \text{ by exercise 3.25} \right).$$

Then $P(t < X \leq t + \delta \mid X > t) \approx h(t)\delta$
for small δ.

Fact: $X \sim \exp(\beta)$ iff $h(t) = \frac{1}{\beta}$
for all t.

The exponential distn. has a constant hazard rate, and it is the only distn. with this property.

(Note: "constant hazard rate" property is equivalent to "memoryless" property.)

Proof of one direction:

Suppose $X \sim \exp(\beta)$. Then

$$h(t) = \lim_{\delta \downarrow 0} \frac{P(t < X \leq t+\delta \mid X > t)}{\delta}$$

$$= \lim_{\delta \downarrow 0} \frac{1}{\delta} \frac{P(t < X \leq t+\delta)}{P(X > t)}$$

$$= \lim_{\delta \downarrow 0} \frac{1}{\delta} \frac{F(t+\delta) - F(t)}{1 - F(t)}$$

$$= \lim_{\delta \downarrow 0} \frac{(1 - e^{-(t+\delta)/\beta}) - (1 - e^{-t/\beta})}{\delta e^{-t/\beta}}$$

$$= \lim_{\delta \downarrow 0} \left(\frac{1 - e^{-\delta/\beta}}{\delta} \right) = \frac{1}{\beta}$$

by L'Hospital's.

Fact: If $Z_1, Z_2, \ldots, Z_n$ are iid $\exp(\beta)$,
then $X \equiv \min_i Z_i \sim \exp(\beta/n)$.

"**Proof**": (heuristic)

X has the memoryless property.

Thus $X \sim \exp(\xi)$ for some value ξ.

What is ξ?

Determine ξ by considering the failure rate.

"Clearly"

$$h_X(t) = n h_{Z_1}(t) = n \cdot \frac{1}{\beta}.$$

But $h_X(t) = \frac{1}{\xi}$.

Thus $\frac{1}{\xi} = \frac{n}{\beta} \Rightarrow \xi = \frac{\beta}{n}$.

A more formal proof uses:

Lemma: If $X_1, X_2, \ldots, X_n$ are independent with cdf's $F_1, F_2, \ldots, F_n$, then

$$F_{\min X_i}(t) = 1 - \prod_{i=1}^{n} (1 - F_i(t)) \quad \text{and}$$

$$F_{\max X_i}(t) = \prod_{i=1}^{n} F_i(t).$$

Consequences of Lemma:

If $X_1, \ldots, X_n$ are i.i.d. with cdf F, then

$$F_{\min X_i}(t) = 1 - (1 - F(t))^n$$

$$F_{\max X_i}(t) = (F(t))^n$$

If $X_1, \ldots, X_n$ are iid $\exp(\beta)$ rv's, then

$$F_{\min X_i}(t) = 1 - (e^{-t/\beta})^n = 1 - e^{-t/(\beta/n)} \quad \text{for } t \geq 0,$$

(Thus $\min X_i \sim \exp(\beta/n)$.)

$$F_{\max X_i}(t) = (1 - e^{-t/\beta})^n \quad \text{for } t \geq 0.$$

Proof of Lemma:

$$F_{\max X_i}(t) = P(\max_{1\le i\le n} X_i \le t)$$

$$= P\left(\bigcap_{i=1}^{n} \{X_i \le t\}\right)$$

(since $\max X_i \le t$ if and only if all the values X_i are $\le t$)

$$= \prod_{i=1}^{n} P(X_i \le t) \quad \text{by independence}$$

$$= \prod_{i=1}^{n} F_i(t)$$

$$F_{\min X_i}(t) = P(\min X_i \le t) = 1 - P(\min X_i > t)$$

$$= 1 - P\left(\bigcap_{i=1}^{n} \{X_i > t\}\right)$$ (since $\min X_i > t$ iff all the values X_i are $> t$)

$$= 1 - \prod_{i=1}^{n} P(X_i > t) = 1 - \prod_{i=1}^{n} (1 - F_i(t))$$

Problem: (The "K stall" problem.)

A room is lit by K light bulbs. The lifetimes of the bulbs are iid $Exp(\beta)$. How long until the room is in total darkness?

Find the mean and variance of the time until total darkness.

<u>Solution</u>: Assume K lightbulbs with lifetimes $Z_1, Z_2, \ldots, Z_K \sim$ iid $\exp(\beta)$.

Define

$$Z_{(1)} = Z_{(1:K)} = \min(Z_1, Z_2, \ldots, Z_K)$$

$$= \text{time at which first bulb burns out},$$

$$Z_{(i)} = Z_{(i:K)} = \text{time at which } i^{\text{th}} \text{ bulb burns out},$$

$$Z_{(K)} = Z_{(K:K)} = \text{time at which last bulb burns out}$$

$$= \text{time until total darkness}$$

$$= \max(Z_1, Z_2, \ldots, Z_K).$$

Note that $Z_{(1)} < Z_{(2)} < \cdots < Z_{(K)}$ with probability one (that is, ties have probability zero).

Consider the identity

$$Z_{(K)} = Z_{(1)} + (Z_{(2)} - Z_{(1)}) + \cdots + (Z_{(K)} - Z_{(K-1)}).$$

The Memoryless Property implies:

If at any time j bulbs remain, the amount of time until the <u>next</u> bulb burns out does not depend on how long the bulbs have been in operation.

At time $Z_{(i:K)}$ there are $K-i$ bulbs remaining. The memoryless property says they are all as good as new. The time until the next bulb burns out is the minimum of $K-i$ exponential rv's which are iid $\exp(\beta)$. Thus

$$Z_{(i+1:K)} - Z_{(i:K)} \sim \exp\left(\frac{\beta}{K-i}\right)$$

and this rv is independent of what has happened already $Z_{(1)}, Z_{(2)}, \ldots, Z_{(i)}$.

Thus (intuitively)

$$Z_{(1)}, \; Z_{(2)} - Z_{(1)}, \ldots, \; Z_{(K)} - Z_{(K-1)}$$

are <u>independent</u> exponential rv's

with $Z_{(1)} \sim \exp\left(\frac{\beta}{K}\right)$ and

$$Z_{(i+1)} - Z_{(i)} \sim \exp\left(\frac{\beta}{K-i}\right).$$

Thus

$$Z_{(K)} = Z_{(1)} + (Z_{(2)} - Z_{(1)}) + \cdots + (Z_{(K)} - Z_{(K-1)})$$

implies

$$\begin{aligned} E Z_{(K)} &= E Z_{(1)} + E(Z_{(2)} - Z_{(1)}) + \cdots \\ &\qquad + E(Z_{(K)} - Z_{(K-1)}) \\ &= \frac{\beta}{K} + \frac{\beta}{K-1} + \frac{\beta}{K-2} + \cdots + \frac{\beta}{2} + \beta \end{aligned}$$

and (by independence)

$$\begin{aligned} \mathrm{Var}\, Z_{(K)} &= \mathrm{Var}(Z_{(1)}) + \mathrm{Var}(Z_{(2)} - Z_{(1)}) \\ &\qquad + \cdots + \mathrm{Var}(Z_{(K)} - Z_{(K-1)}) \\ &= \left(\frac{\beta}{K}\right)^2 + \left(\frac{\beta}{K-1}\right)^2 + \cdots + \left(\frac{\beta}{2}\right)^2 + \beta^2. \end{aligned}$$

$$F_{\underbrace{\max Z_i}_{X}}(t) = (1-e^{-t/\beta})^K.$$

$$\text{pdf} \quad n(1-e^{-t/\beta})^{K-1}\left(\tfrac{1}{\beta}e^{-t/\beta}\right)$$

$$EX = \int_0^\infty (1 - F_X(t))\,dt$$

for any nonnegative r.v. Z

$$= \int_0^\infty (1-(1-e^{-t/\beta})^K)\,dt$$

setting $K=3$

$$= \int_0^\infty (1-(1-3e^{-t/\beta}+3e^{-2t/\beta}-e^{-3t/\beta}))\,dt$$

$$= 3\beta - 3\left(\frac{\beta}{2}\right) + \frac{\beta}{3}$$

$$= \beta + \frac{\beta}{2} + \frac{\beta}{3}$$ agrees with previous.

Can also obtain variance in similar fashion.

A Connection between the Gamma distn. and the Poisson distn.
(via the Poisson process)

Fact: A process of arrivals is a Poison process with rate λ
if and only if
the interarrival times are i.i.d. exponential rv's with mean $\frac{1}{\lambda}$.

Notation: Let (time starts from zero)

T_r = time of r^{th} arrival,

S_t = # of arrivals during $(0,t)$.

The interarrival times are

$$T_1, T_2 - T_1, T_3 - T_2, \ldots, T_r - T_{r-1}, \ldots$$

Note that

$T_r \sim \text{Gamma}(r, \frac{1}{\lambda})$ (by fact above)

$S_t \sim \text{Poisson}(\lambda t)$ (by defn. of Poisson process)

Thus

$$\{T_r > t\} = \{S_t < r\}$$

implies $P(T_r > t) = P(S_t < r)$

which leads to

$$\int_t^\infty \underbrace{\frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}}_{\substack{\text{Gamma}(\alpha = r,\, \beta = \frac{1}{\lambda}) \\ \text{pdf}}} dx = \sum_{i=0}^{r-1} \underbrace{\frac{(\lambda t)^i e^{-\lambda t}}{i!}}_{\substack{\text{Poisson}(\lambda t) \\ \text{pmf}}} .$$

This formula can also be derived by repeated integration by parts.

Fact: For a Poisson process with rate λ, the time until the first arrival has an exponential distribution with mean

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

Proof: (Start time from zero.)

Let T_1 = the time of the first arrival,

S_t = the # of arrivals by time t.

We know $S_t \sim \text{Poisson}(\lambda t)$ so that

$$P(S_t = 0) = \left.\frac{(\lambda t)^k e^{-\lambda t}}{k!}\right|_{k=0} = e^{-\lambda t}.$$

But $\{T_1 > t\} = \{S_t = 0\}$ (Think about it!)

Thus $P(T_1 > t) = P(S_t = 0) = e^{-\lambda t}$

and $P(T_1 \leq t) = 1 - e^{-\lambda t}$.

This is the cdf of $\text{Exp}(1/\lambda)$ distn.

Central Limit Theorem (CLT)

If $X_1, X_2, \ldots, X_n$ are iid from a distribution with $E X_i = \mu$ and $\operatorname{Var} X_i = \sigma^2 < \infty$,

then
$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0,1) \quad (\text{as } n \to \infty)$$

and, equivalently,
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \quad (\text{as } n \to \infty)$$

where $S_n = \sum_{i=1}^{n} X_i$ and $\bar{X}_n = S_n/n$.

Informal Statement

For "large" n,

$$S_n \sim \text{approx } N(n\mu, n\sigma^2),$$
$$\bar{X}_n \sim \text{approx } N\left(\mu, \frac{\sigma^2}{n}\right).$$

How large should n be?

Normal Approximations

Binomial distn.

If $X \sim \text{Binomial}(n,p)$ and n is large, then

$$X \sim \text{approx Normal}(\mu = np, \sigma^2 = np(1-p)).$$

Argument: $X = \sum_{i=1}^{n} Z_i$ where

$Z_1, Z_2, \ldots, Z_n$ are iid Bernoulli(p).

Now use CLT.

Poisson distn.

If $X \sim \text{Poisson}(\lambda)$ and λ is large,

then $X \sim \text{approx Normal}(\mu = \lambda, \sigma^2 = \lambda)$.

Argument: Fix a value λ_0.

Let $Z_1, Z_2, \ldots, Z_n$ be iid Poisson(λ_0).

Then $\sum_{i=1}^{n} Z_i \sim \text{Poisson}(\lambda = n\lambda_0)$.

Now let $n \to \infty$ and use CLT.

Negative Binomial distn

If $X \sim NegBin(r, p)$ and r is large,
then $X \sim$ approx $N\left(\mu = \frac{r}{p}, \sigma^2 = r \cdot \frac{1-p}{p^2}\right)$

Argument: $X = \sum_{i=1}^{r} Z_i$ where
$Z_1, Z_2, \ldots, Z_r$ are iid Geometric(p)
Now use CLT.

Gamma distn.

If $X \sim Gamma(\alpha, \beta)$ and α is large,
then X is approximately Normal.

Argument: Fix a value α_0.
Let $Z_1, Z_2, \ldots, Z_n$ be iid Gamma(α_0, β).
Then $\sum_{i=1}^{n} Z_i \sim Gamma(n\alpha_0, \beta)$.
Now let $n \to \infty$ and use CLT.

Beta distn

If $X \sim Beta(\alpha, \beta)$ with both α, β large,
then X is approx. Normal.

and others...

Comments on C5

Suppose

n people support candidate A.

n people support B.

Each person votes with probability p.

People decide independently.

Let X = # who vote for A,

Y = # who vote for B.

Then X and Y are independent with a Binomial(n, p) distribution

which is approximately

Normal$(\mu = np, \sigma^2 = np(1-p))$

if n is sufficiently large.

$P(\text{tie}) = P(X = Y) = P(X - Y = 0)$.

Recall: If X, Y independent with

$X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$,

then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

and $X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.

Thus, for n large,

$$D \equiv X - Y \sim N(\mu - \mu, \sigma^2 + \sigma^2) \quad \text{(approximately)}$$

$$\sim N(0, 2\sigma^2)$$

Let $D^* \sim N(0, 2\sigma^2)$.

$\left(\begin{array}{l} D \text{ is approx } N(0, 2\sigma^2). \\ D^* \text{ is exactly } N(0, 2\sigma^2) \end{array} \right)$

Since D is integer-valued, the continuity correction gives

$$P(D=0) \approx P\left(-\tfrac{1}{2} < D^* < \tfrac{1}{2}\right)$$

$$= P\left(\frac{-\frac{1}{2} - 0}{\sqrt{2\sigma^2}} < \frac{D^* - 0}{\sqrt{2\sigma^2}} < \frac{\frac{1}{2} - 0}{\sqrt{2\sigma^2}} \right)$$

$$= P(-\varepsilon < Z < \varepsilon)$$

where $Z = \dfrac{D^* - 0}{\sqrt{2\sigma^2}} \sim N(0,1)$

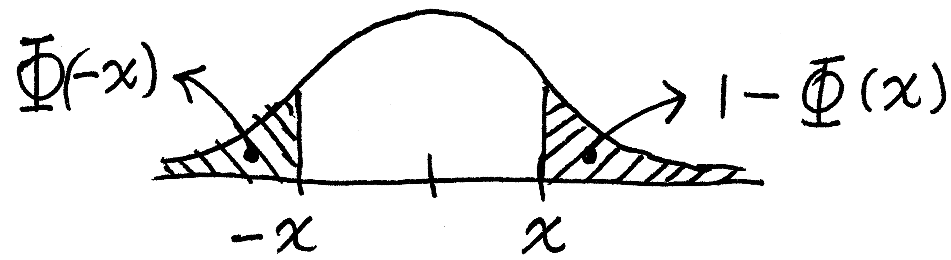
and $\varepsilon = (\tfrac{1}{2} - 0)/\sqrt{2\sigma^2}$.

Note: If n is large, $\sigma^2 = np(1-p)$ is also large so that ε will be small.

$$= \Phi(\varepsilon) - \Phi(-\varepsilon) \quad \text{where } \Phi \text{ is the cdf of } N(0,1)$$

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

by symmetry of the N(0,1) distn.



$$\Phi(-x) = 1 - \Phi(x) \text{ for all } x.$$

Thus, to compute P(tie) you calculate $\varepsilon = \dfrac{1}{2\sqrt{2np(1-p)}}$

and use tables or computer to get

$$2\Phi(\varepsilon)-1.$$

<u>Problem with Tables</u>

Tables typically give $\Phi(z)$ or $P(Z>z)$ for $z = 0.00, 0.01, 0.02, \ldots, 3.49$ (say).

If ε is small, rounding to the nearest tabled z-value loses a lot of accuracy.

Better to use ...

Good approximation for small ϵ

Let $\varphi(z) = \dfrac{e^{-z^2/2}}{\sqrt{2\pi}}$ = pdf for $N(0,1)$.

Then $P(\text{tie})$ is approximately

$$P(-\epsilon < \underset{\substack{\uparrow \\ N(0,1)}}{Z} < \epsilon) = \int_{-\epsilon}^{\epsilon} \varphi(z)\,dz$$

$$\approx (2\epsilon)\,\varphi(0) = \frac{2\epsilon}{\sqrt{2\pi}}$$ when ϵ is small.

Thus $P(\text{tie}) \approx \dfrac{2}{\sqrt{2\pi}} \cdot \dfrac{1}{2\sqrt{2np(1-p)}}$

$$= \frac{1}{\sqrt{4\pi n p(1-p)}}.$$

When $n = 2{,}000{,}000$, and $p = 0.8$
you get
$P(\text{tie}) \approx .0004987$

Note that $\epsilon = .000625$ which rounds to zero giving $2\underbrace{\Phi(0)}_{.5} - 1 = 0$.

Problem actually wants

P(Joe makes a difference)

= P(Joe breaks a tie) + P(Joe creates a tie)

$= P(D=0) + P(D=1)$

(assuming Joe votes for B)

$= P(0 \leq D \leq 1)$

$\approx P(-.5 \leq D^* \leq 1.5)$

$$= P\left(\frac{-.5}{\sqrt{2\sigma^2}} < Z < \frac{1.5}{\sqrt{2\sigma^2}}\right)$$

$$\approx \underbrace{\left[\frac{1.5}{\sqrt{2\sigma^2}} - \left(\frac{-.5}{\sqrt{2\sigma^2}}\right)\right]}_{\text{width of interval of integration}} \cdot \underbrace{\frac{1}{\sqrt{2\pi}}}_{\text{density of } Z \text{ at zero}} \quad \left(\begin{array}{l}\text{when } n \\ \text{is large}\end{array}\right)$$

$$= \frac{2}{\sqrt{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}} = \frac{1}{\sqrt{\pi\sigma^2}}$$

$$= \frac{1}{\sqrt{\pi n p(1-p)}} = .000997$$

when $n = 2000000$, $p = .8$.

Beta Distribution

$X \sim \text{Beta}(\alpha, \beta)$ with $\alpha > 0, \beta > 0$
has pdf

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad \text{for } 0<x<1$$

where $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$

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

Parameters α, β are "shape" parameters.

$$\alpha<1 \Rightarrow \lim_{x\to 0^+} f(x) = \infty$$

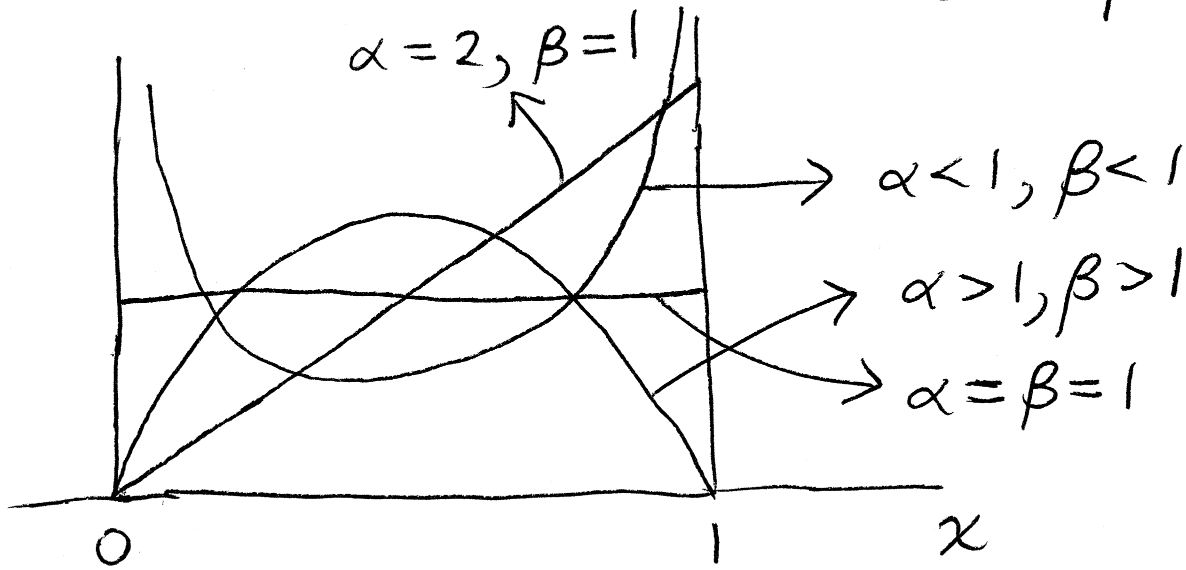
$$\alpha>1 \Rightarrow \lim_{x\to 0^+} f(x) = 0$$

$$\beta<1 \Rightarrow \lim_{x\to 1^-} f(x) = \infty$$

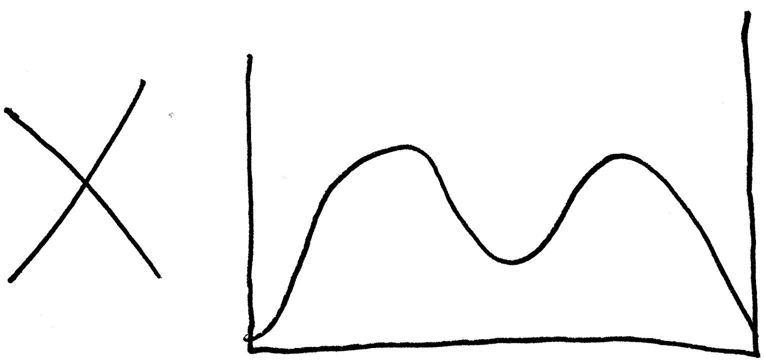
$$\beta>1 \Rightarrow \lim_{x\to 1^-} f(x) = 0$$

Setting $\alpha=\beta=1$ gives Uniform$(0,1)$ distn.

Modifying α and β gives a variety of shapes.



But you can**not** get a bimodal distribution



(unless the pdf blows up at 0 and 1 giving modes at 0 and 1)

The Beta distribution is used to model quantities taking values in $(0,1)$ (such as proportions and probabilities).

For a Beta distribution:

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

$$\mathrm{Var}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

which are easy consequences of

$$EX^k = \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+k)}$$

$$= \frac{(\alpha+k-1)(\alpha+k-2)\cdots\alpha}{(\alpha+\beta+k-1)\cdots(\alpha+\beta)}$$

MGF $M_X(t)$ exists for all t,
(since X is bounded)
but is not useful. (It does not have a simple closed form.)