Lesson 3: Probability Distributions

(a bestiary)

Review

all the same properties apply to random variables

$$Pr(A,B)$$ joint distribution

$$Pr(A|B) = \frac{Pr(A,B)}{Pr(B)}$$ conditional distribution

$$Pr(A) = \sum Pr(A|B_i)\,Pr(B_i)$$ marginal distribution

$$Pr(A|B) = \frac{Pr(B|A)\,Pr(A)}{Pr(B)}$$ Baye's Rule

Review

	Discrete	Continuous
PDF	$f(z)=Pr(Z=z)=p_z$ where $\sum(z)=1$ $0\leq p_z\leq 1$	$f(z)=\frac{dF}{dz}$ where $\int(z)=1$ $0\leq p_z$
CDF	$F(z)=Pr(Z\leq z)=\sum_0^z p_z$ where $0\leq F(z)\leq 1$	$F(z)=Pr(Z\leq z)=\int_{-\infty}^{z} f(z)$ where $0\leq F(z)\leq 1$

Moments of probability distributions

$$E[x^n] = \int x^n \cdot f(x)\, dx$$

E[ ] = Expected value

First moment (n=1) = mean

Example: exponential

$$E[x] = \int x \cdot f(x)\, dx = \int_0^\infty x \lambda \exp(-\lambda x)$$

$$= -x \exp(-\lambda x)\Big|_0^\infty + \frac{1}{\lambda} \int_0^\infty \lambda \exp(-\lambda x)$$

$$E[x] = 1/\lambda$$

Properties of means

$E[c] = c$

$E[x + c] = E[x] + c$

$E[cx] = c\ E[x]$

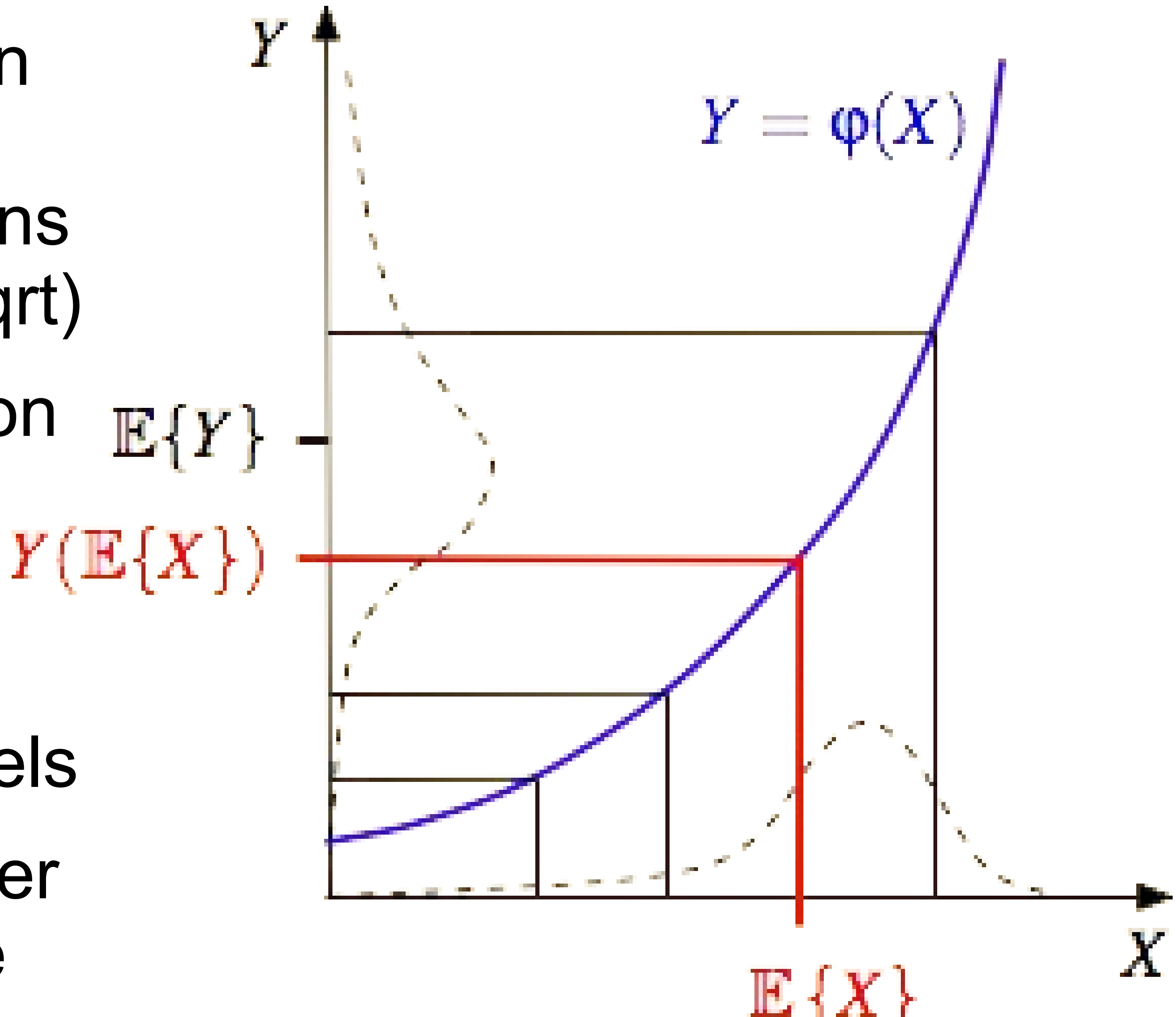
$E[x+y] = E[x] + E[y]$
(even if X is not independent of Y)

$E(xy) = E[x]E[Y]$ **only** if independent

$E[g(x)] != g(E[x])$ **Jensen's Inequality**

Jensen's Inequality

- Rampant misapplication of data transformations (e.g. log or sqrt)
- Transformation of probability distributions
- Nonlinear process models
- Averaging over space or time



Central Moments

$$E\left[\left(x-E[x]\right)^n\right]=\int \left(x-E[x]\right)^n \cdot f(x)\,dx$$

Second Central Moment = Variance = σ^2

Properties of variance

$$Var(aX) = a^2 Var(X)$$

$$Var(X+b) = Var(X)$$

$$Var(X+Y) = Var(X) + Var(Y) + 2\mathrm{Cov}(X,Y)$$

$$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2\mathrm{abCov}(X,Y)$$

$$Var\left(\sum X\right) = \sum Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

$$Var(X) = Var(E[X|Y]) + E[Var(X|Y)]$$

Looking forward...

Use probability distributions to:

- Quantify the match between models and data
- Represent uncertainty about model parameters
- Partition sources of *process* variability

Common Probability Distributions

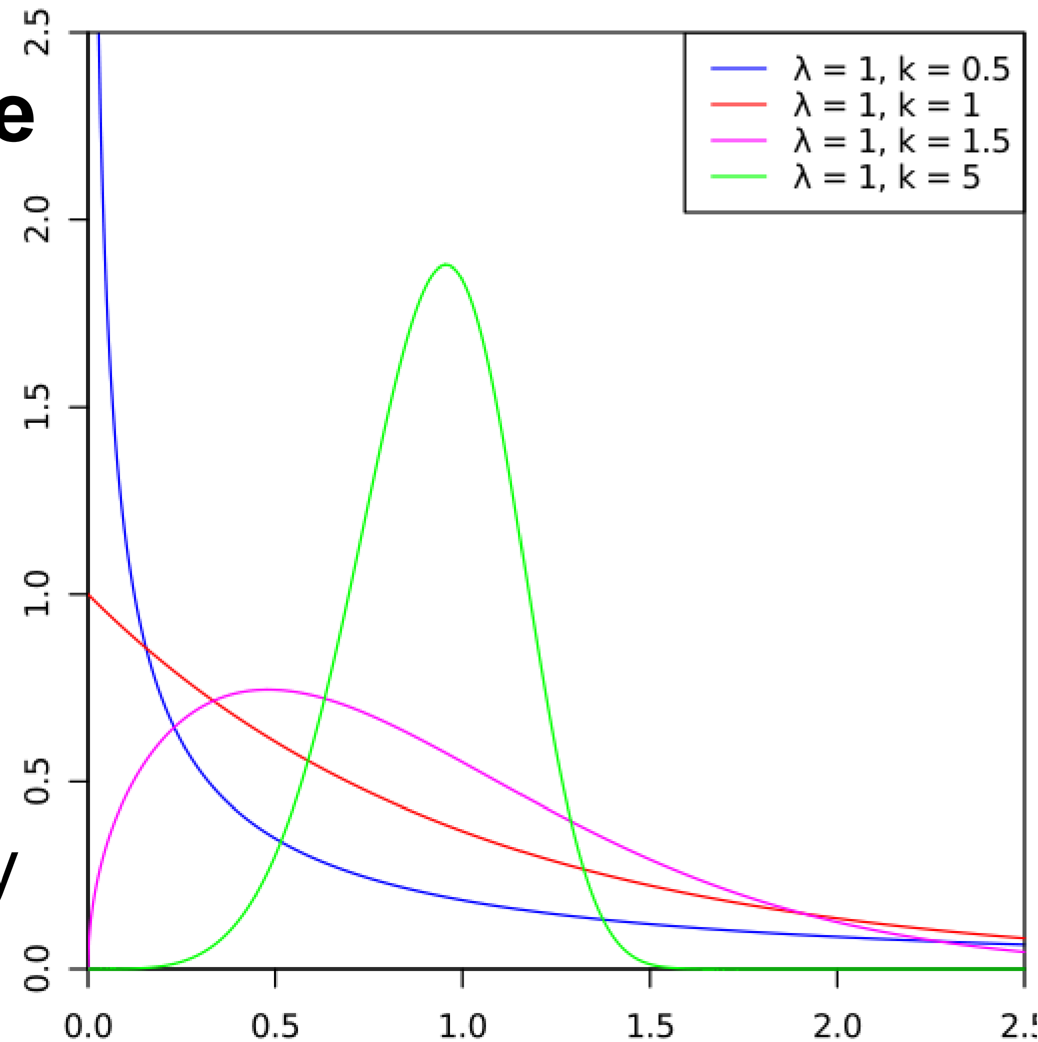
CHARACTERS

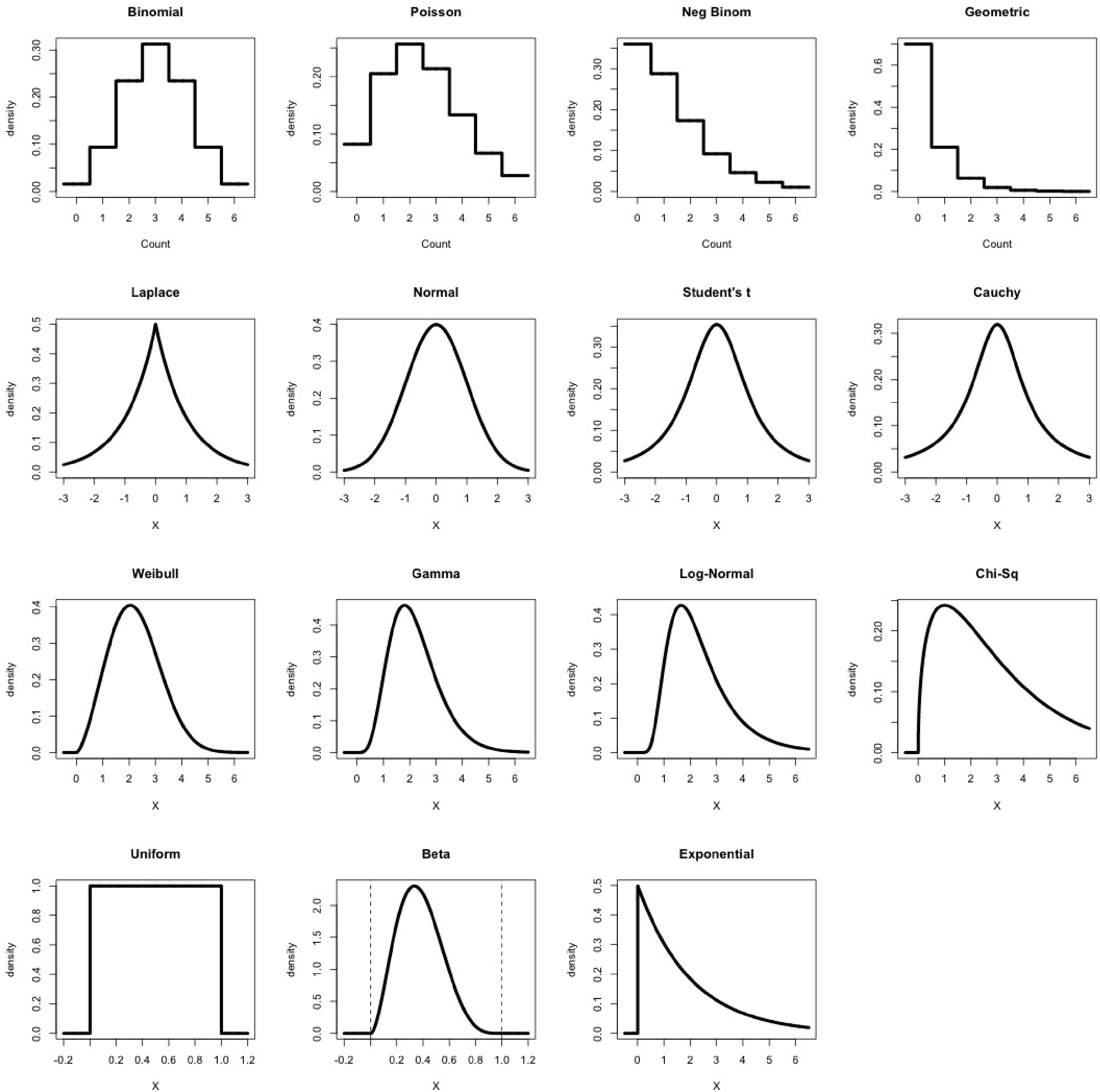
(in order of appearance)

THOMASINA COVERLY, aged thirteen, later sixteen
SEPTIMUS HODGE, her tutor, aged twenty-two, later tw
JELLABY, a butler, middle-aged
EZRA CHATER, a poet, aged thirty-one
RICHARD NOAKES, a landscape architect, middle-aged
LADY CROOM, middle thirties
CAPT. BRICE, RN, middle thirties
HANNAH JARVIS, an author, late thirties
CHLOË COVERLY, aged eighteen
BERNARD NIGHTINGALE, a don, late thirties
VALENTINE COVERLY, aged twenty-five to thirty
GUS COVERLY, aged fifteen
AUGUSTUS COVERLY, aged fifteen

Important Characteristics of PDFs

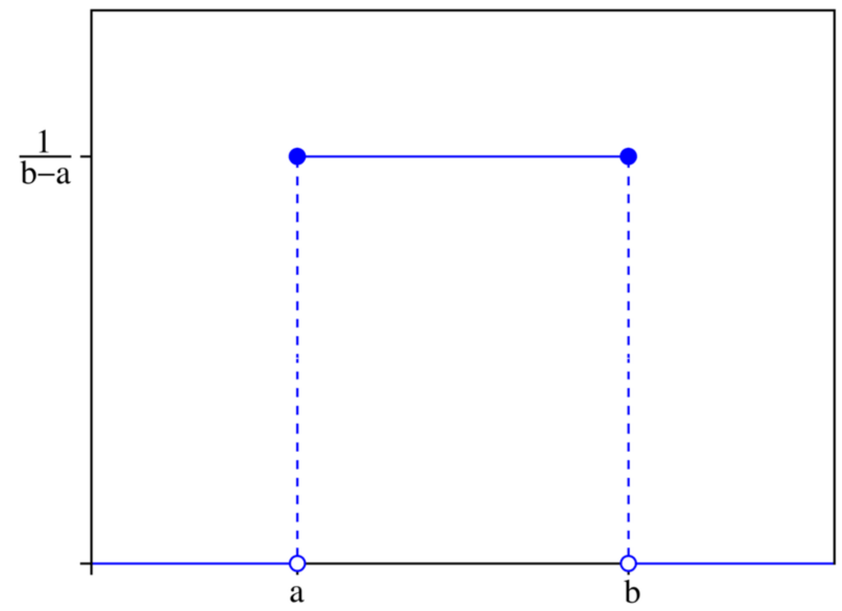
- **Continuous vs discrete**
- **Range restrictions**
- Interpretation
  - Of distribution
  - Of parameters
- Number of parameters
  - Simplicity vs flexibility
- Skew (symmetry)
- Kurtosis (how fat are the tails)

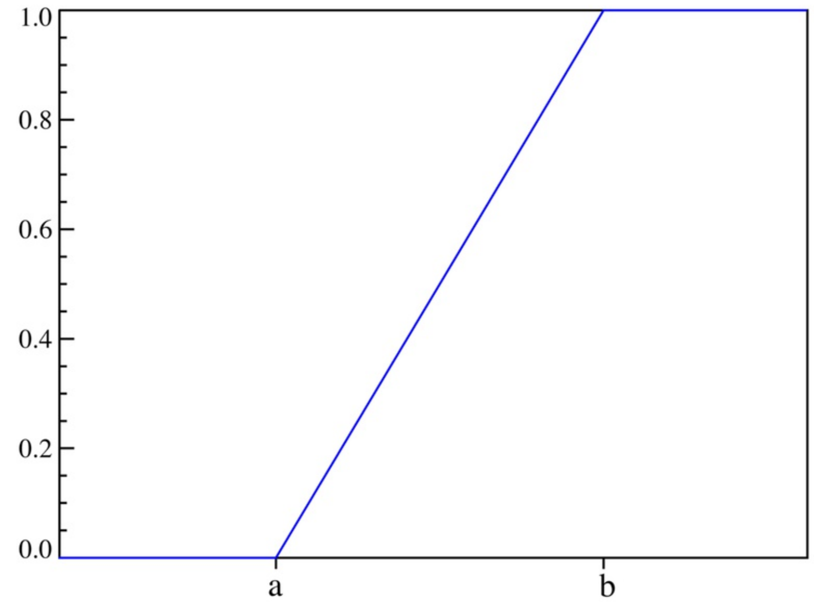



Binomial
Poisson
Neg Binom
Geometric
Count
density
Laplace
Normal
Student's t
Cauchy
Weibull
Gamma
Log-Normal
Chi-Sq
Uniform
Beta
Exponential
X

Uniform

$$Unif(x|a,b) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & x < a \text{ or } x > b \end{cases}$$



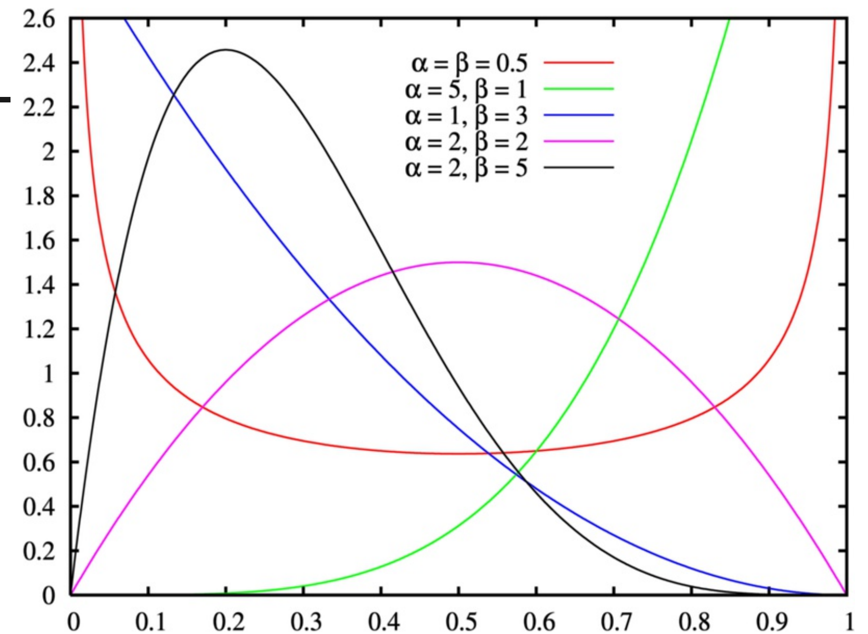


$$E[X] = \frac{a+b}{2}$$

$$Var(X) = \frac{1}{12}(b-a)^2$$

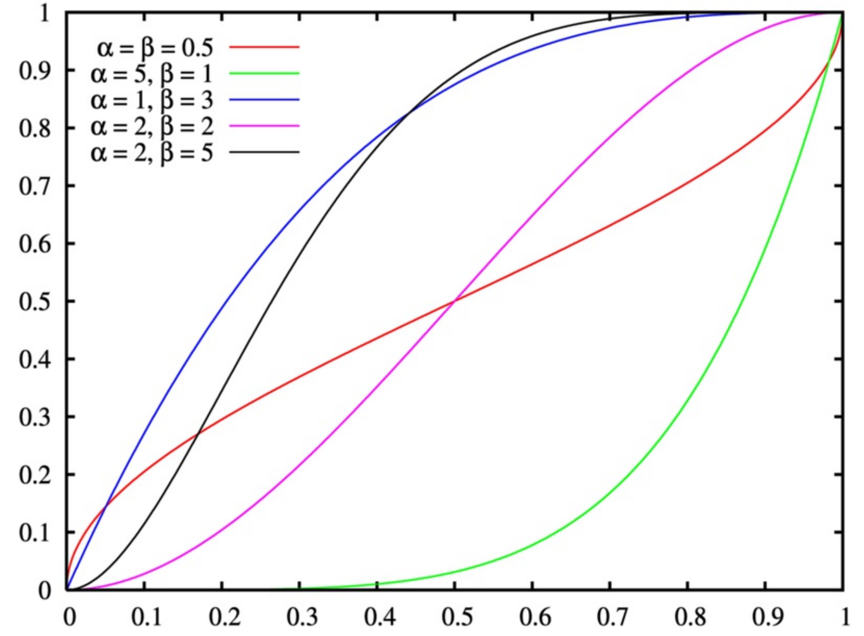
Beta

$$Beta(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$



$$E[X] = \frac{\alpha}{\alpha+\beta}$$

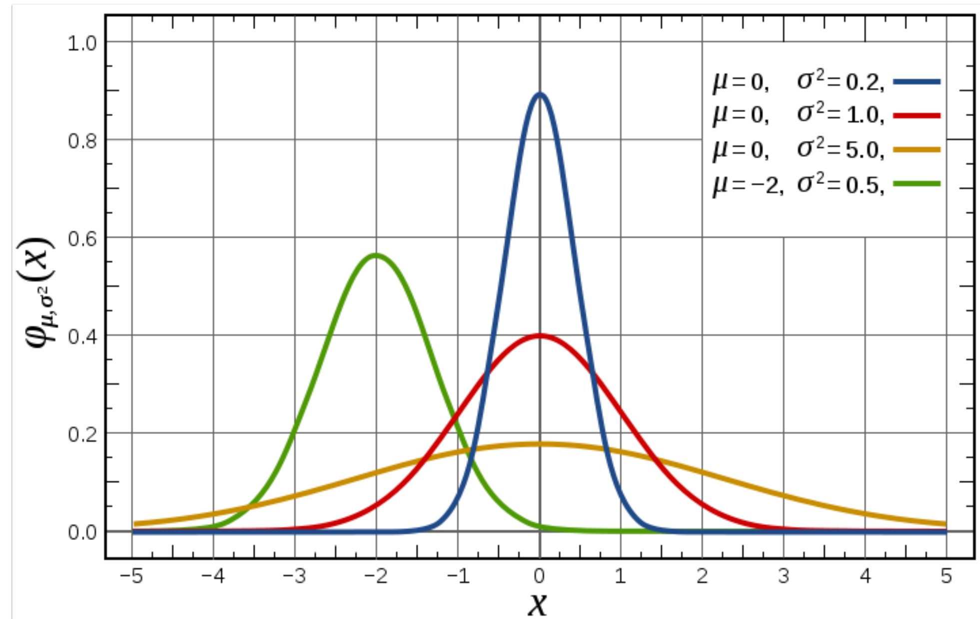
$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$



Defined on (0,1)
Beta(1,1) = Unif(0,1)
Beta(a,b) is symmetric to Beta(b,a)
Multivariate = Dirichlet

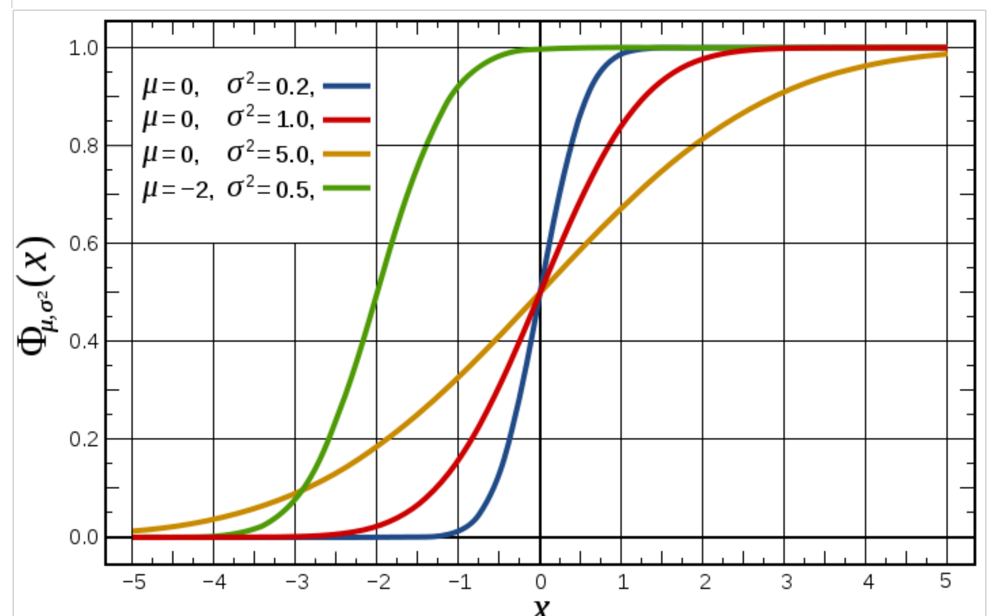
Normal (= Gaussian)

$$N(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$



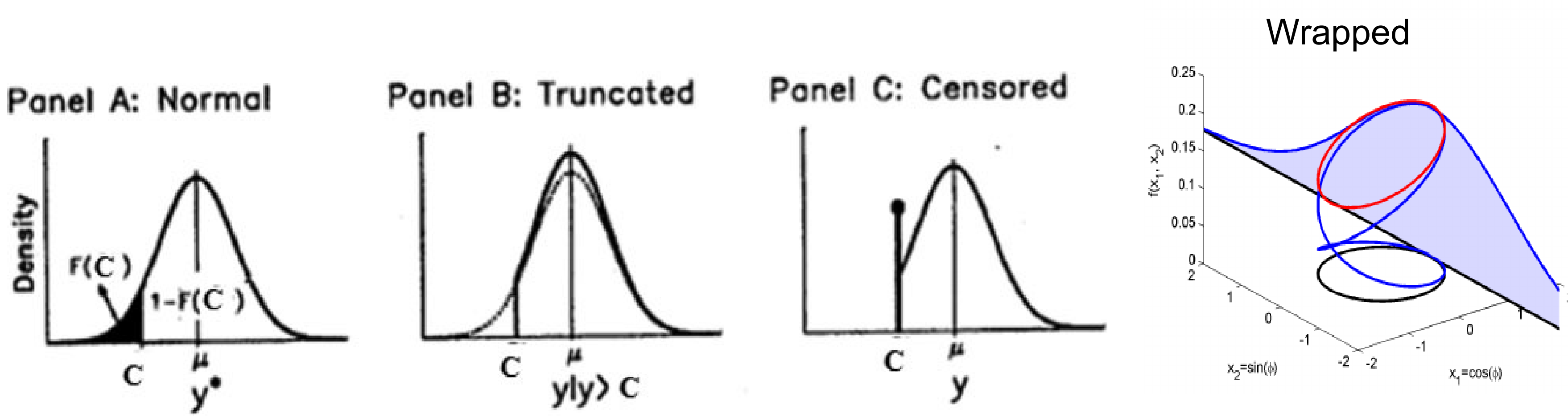
$$E[X] = \mu$$

$$Var(X) = \sigma^2$$



Normal w/ boundaries

- Truncated Normal
- Tobit / Rectified / Censored Normal
- Wrapped Normal



$$E[x] \neq \mu$$

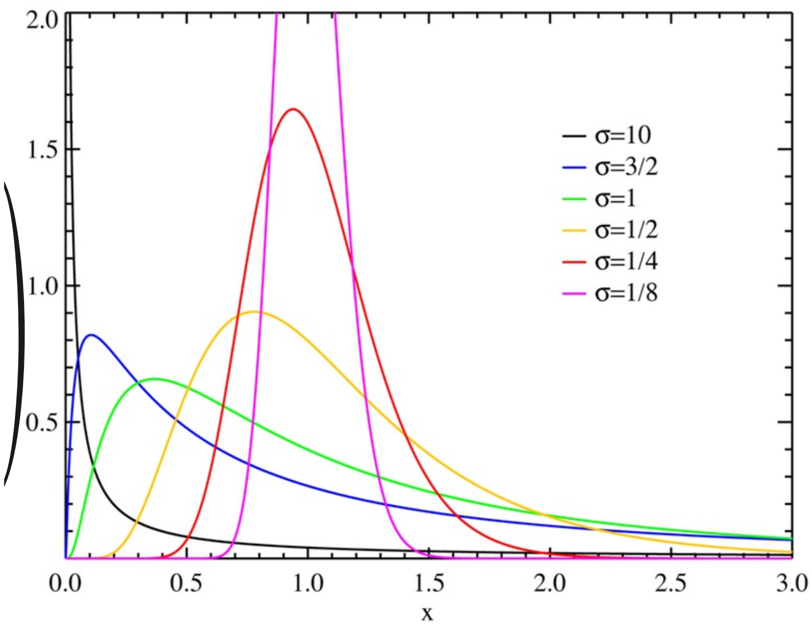
Lognormal

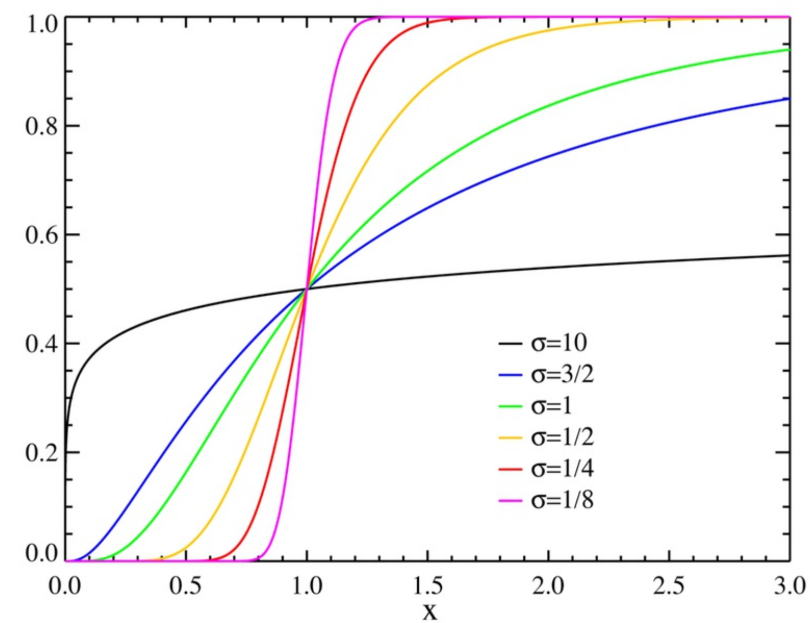
$$\log N(x|\mu,\sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

$$E[X] = e^{\mu+\sigma^2/2}$$

$$Var(X) = \left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$$

μ=0





Exponential

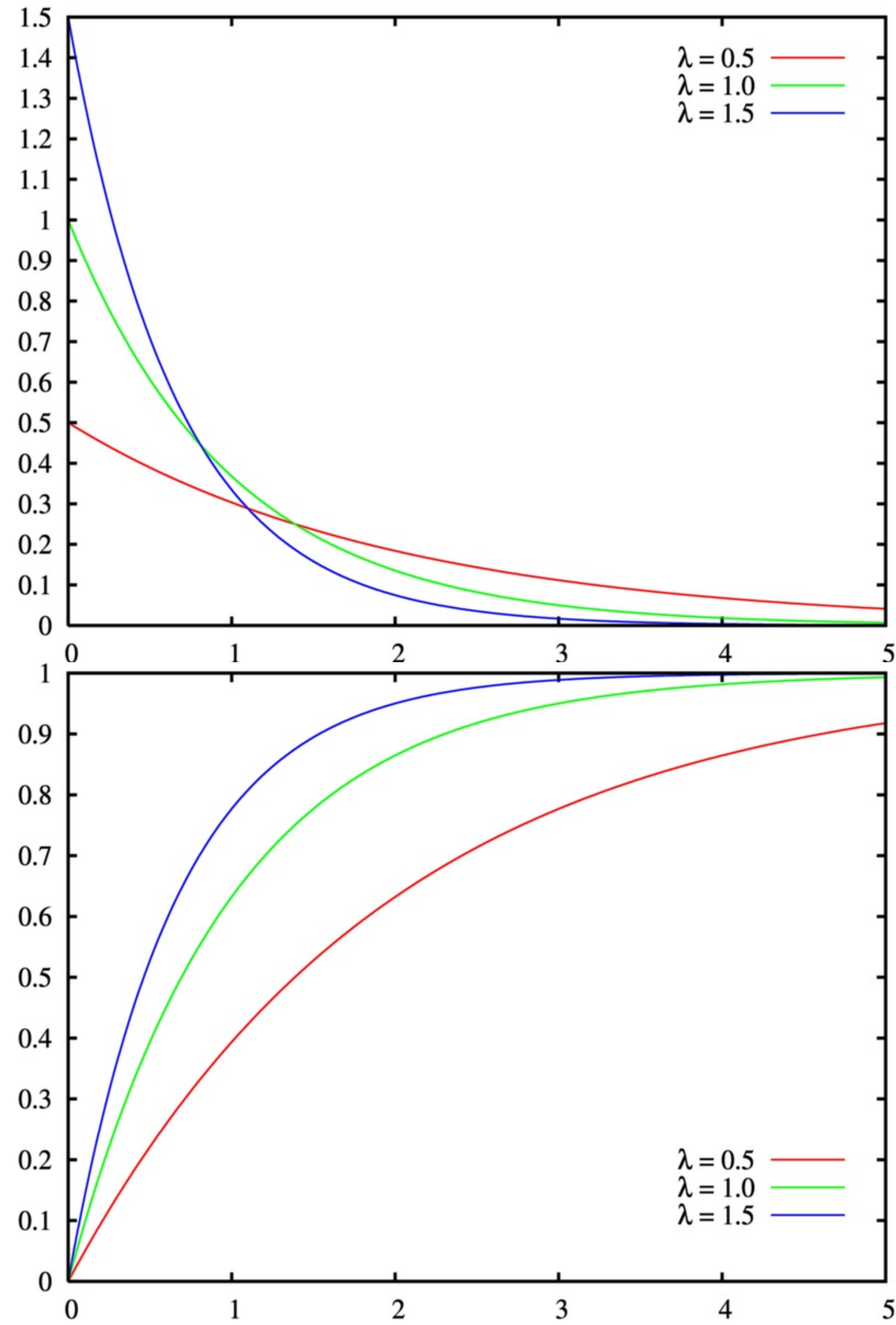
$$\mathrm{Exp}(x|\lambda) = \lambda \exp(-\lambda x)$$

$$E[X] = 1/\lambda$$
$$Var(X) = 1/\lambda^2$$

Distribution of waiting times/distances for a single event

Continuous analog to geometric distrib.

Special case of Gamma(1,λ)



Laplace

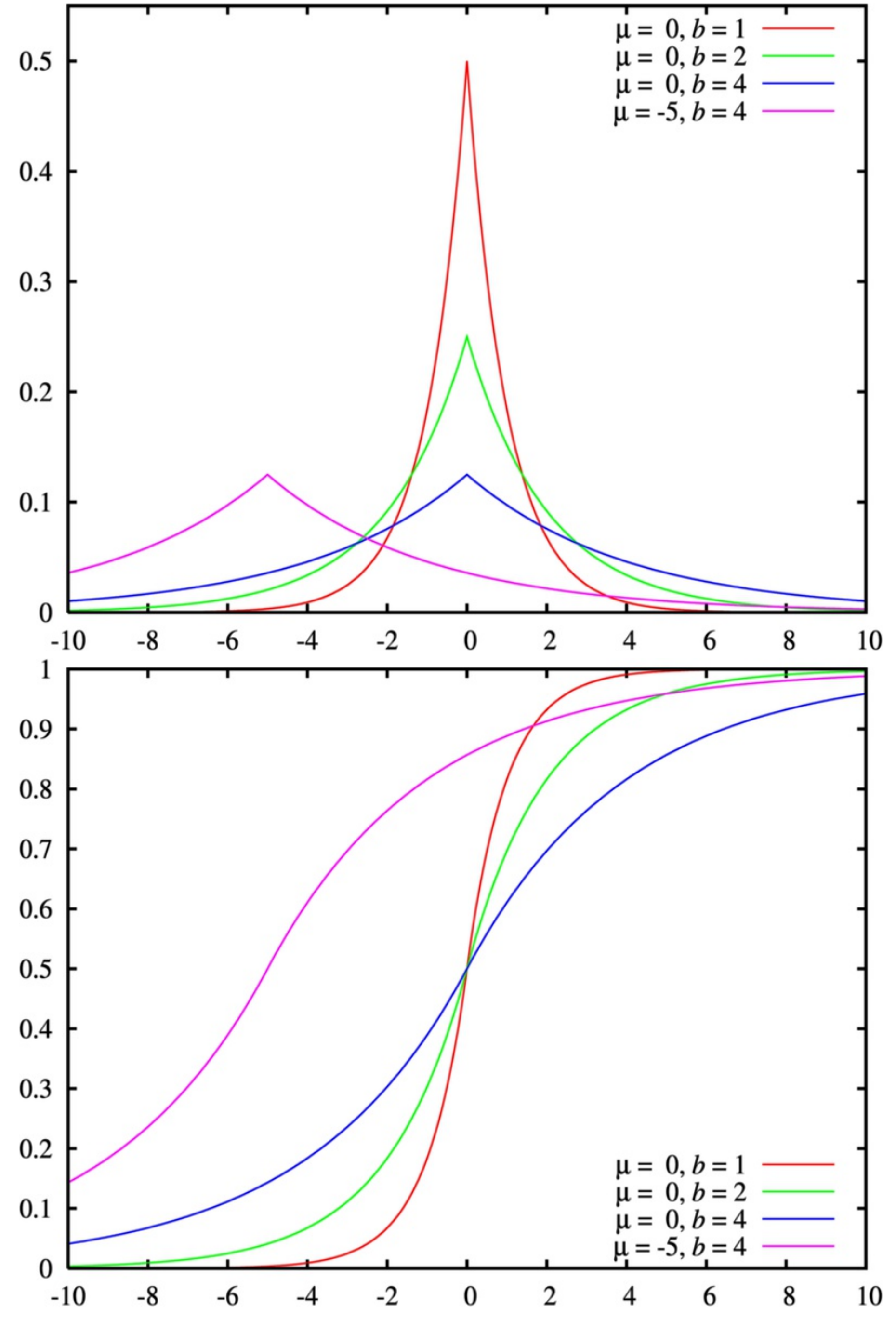
$$\text{Laplace}(x|\mu, b) = \frac{1}{2b}\exp\left(\frac{-|x-\mu|}{b}\right)$$

$$E[X] = \mu$$

$$Var(X) = 2b^2$$

Used to model absolute deviations rather than standard deviations

Fat - tailed



Gamma

$$Gamma(x|a,s) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-x/s}$$

$$E[X] = as = a/r$$

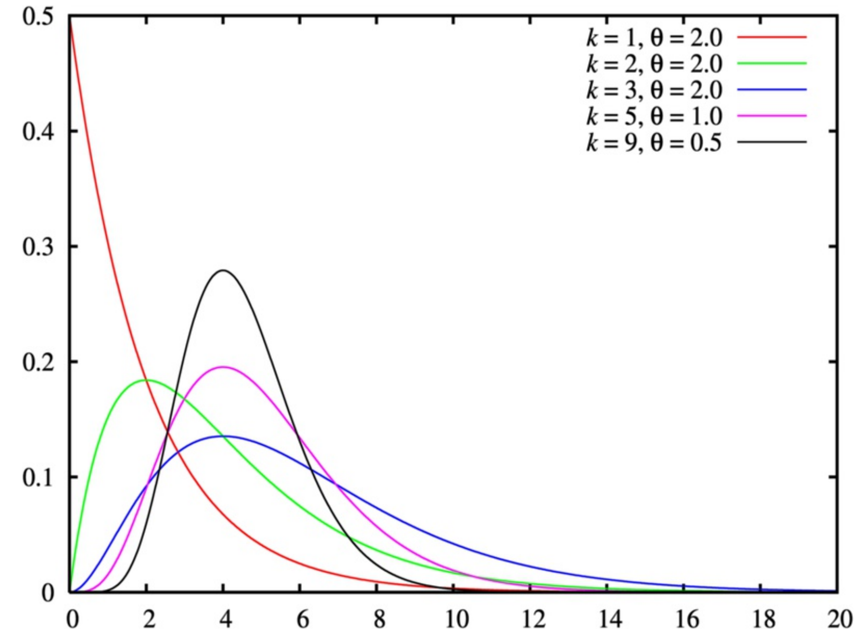
$$Var(X) = as^2 = a/r^2$$

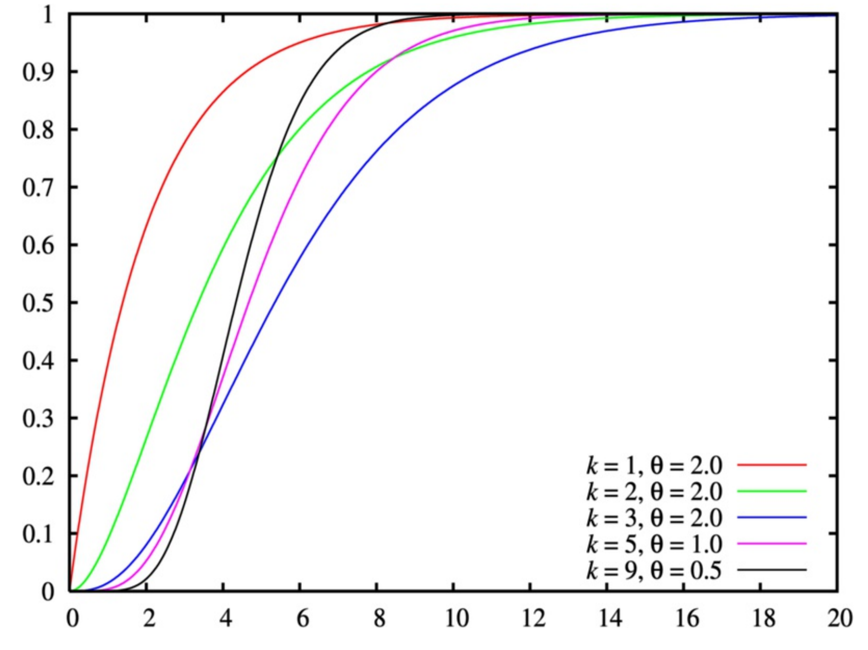
Common alternate parameterization: r = 1/s
a = “shape”, r = “rate”, s = “scale”

Distribution of waiting times until a certain number of events, a, takes place given a rate r

Commonly used to model variances

Multivariate = Wishart





Combinatorics

$$Factorial = n! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \cdots \cdot n$$

$$\Gamma(n) = (n-1)! = \int_0^\infty t^{n-1} e^{-t}\, dt$$

$$\text{Choose} = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$\begin{array}{ccccccccccc} &&&&& 1 &&&&& \\ &&&& 1 && 1 &&&& \\ &&& 1 && 2 && 1 &&& \\ && 1 && 3 && 3 && 1 && \\ & 1 && 4 && 6 && 4 && 1 & \\ 1 && 5 && 10 && 10 && 5 && 1 \end{array}$$

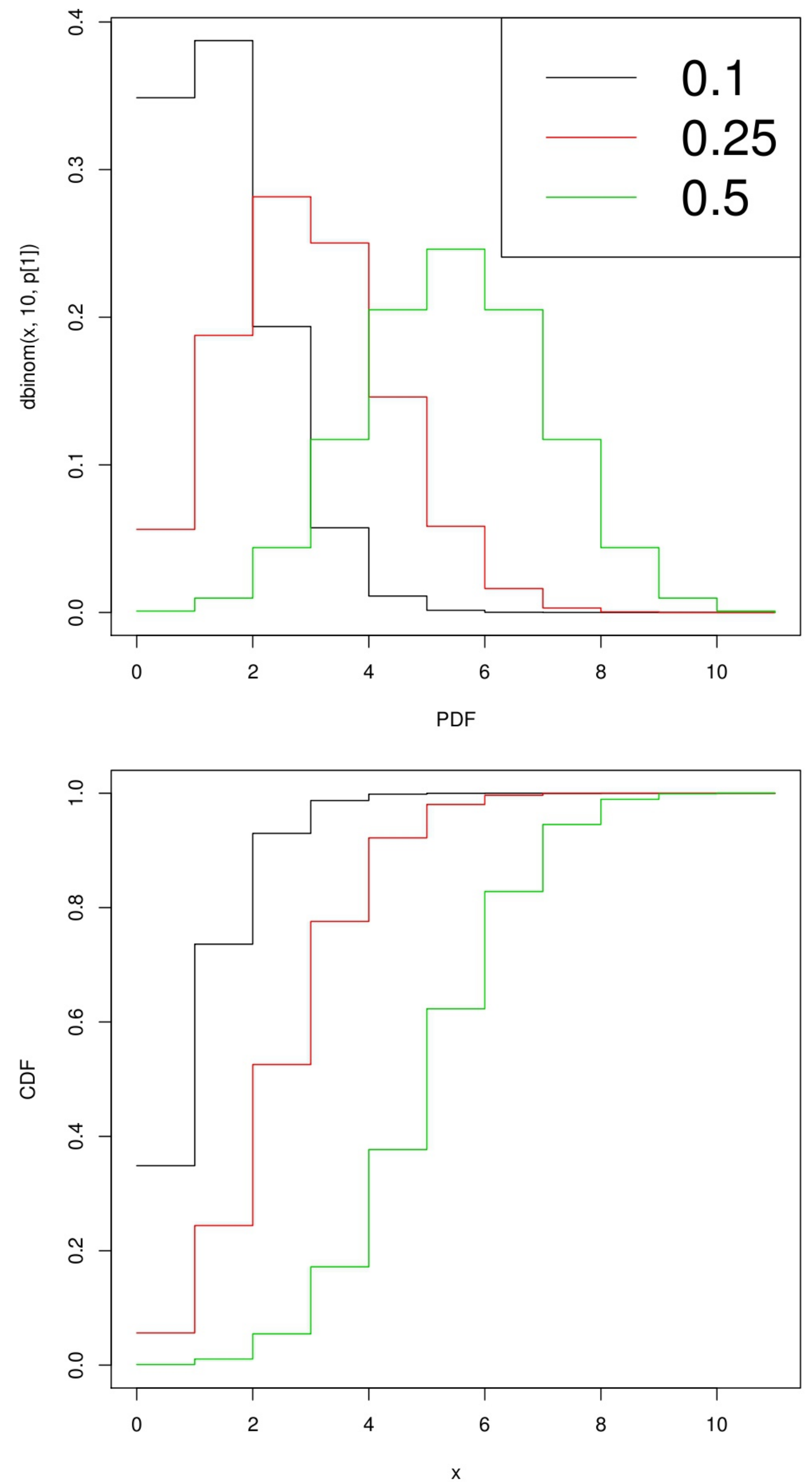
Binomial

$$\mathrm{Binom}(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E[X] = np$$
$$Var(X) = np(1-p)$$

Distribution of the number of “successes”, x, given a sample size n and probability p

Bernoulli distribution is special case of n=1

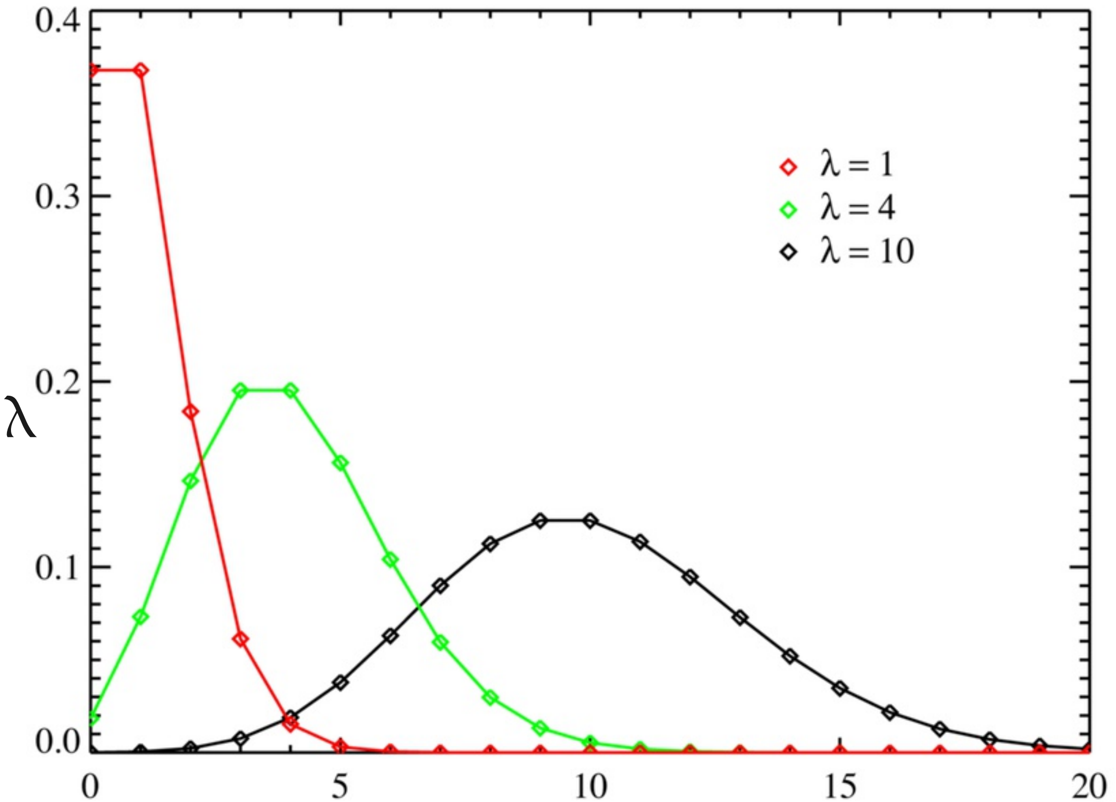


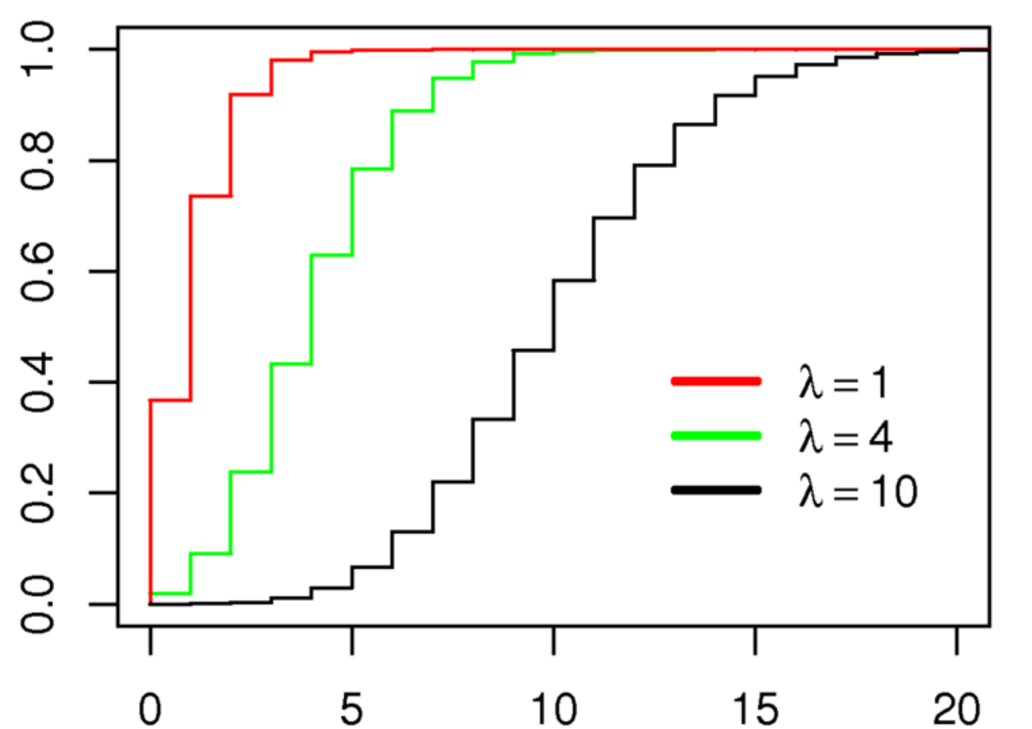
Poisson

$$\text{Poisson}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$E[X] = \lambda$$
$$Var(X) = \lambda$$

number of events that occur in a fixed amount of time or space





Negative Binomial

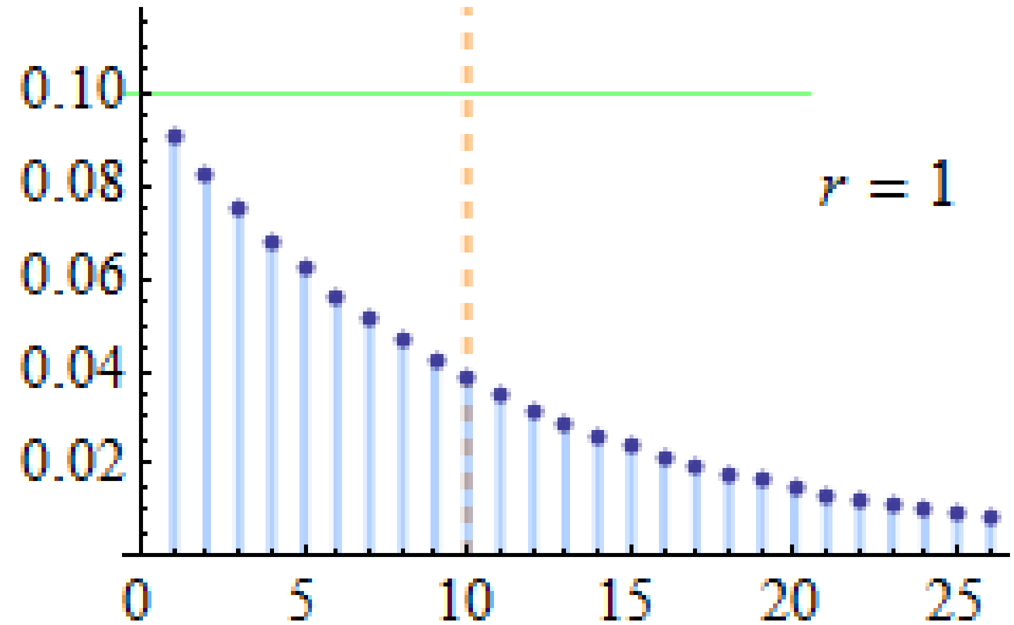
$$\mathrm{NB}(x|n,p) = \binom{x+n-1}{n-1}(1-p)^x p^n$$

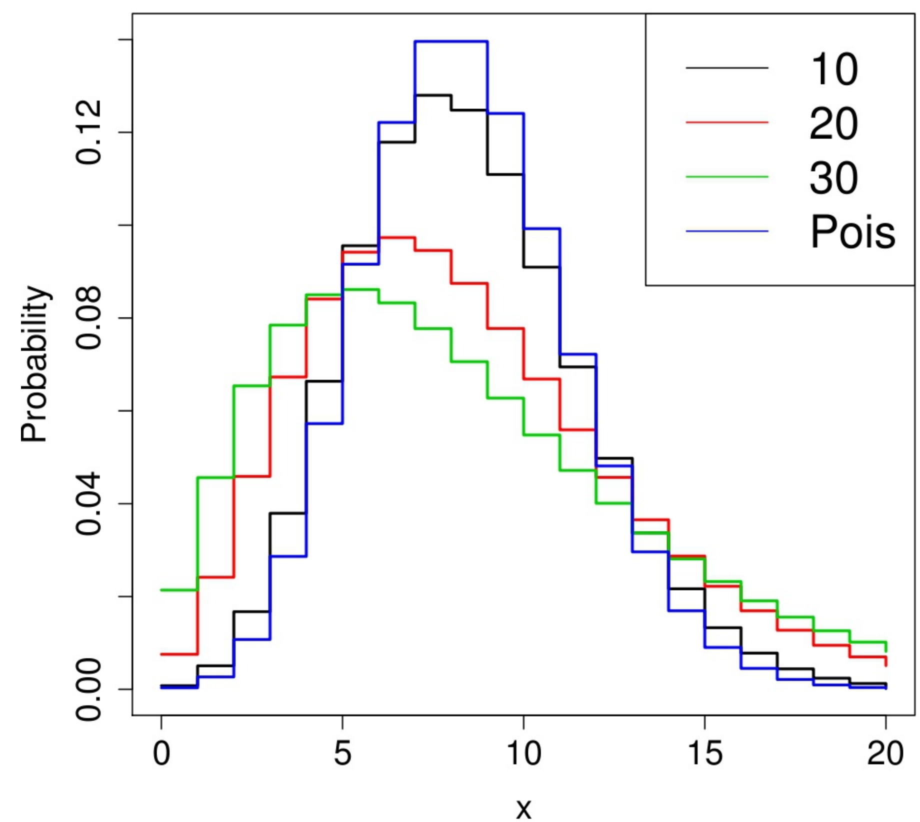
$$E[X] = n\frac{1-p}{p}$$

$$Var(X) = n\frac{1-p}{p^2}$$

n = number of “successes” = r
p = probability of “success”
x = number of “failures” = k
n + x = total number of “trials”

A number of alternative parameterizations, most important is Poisson-Gamma mixture





Geometric

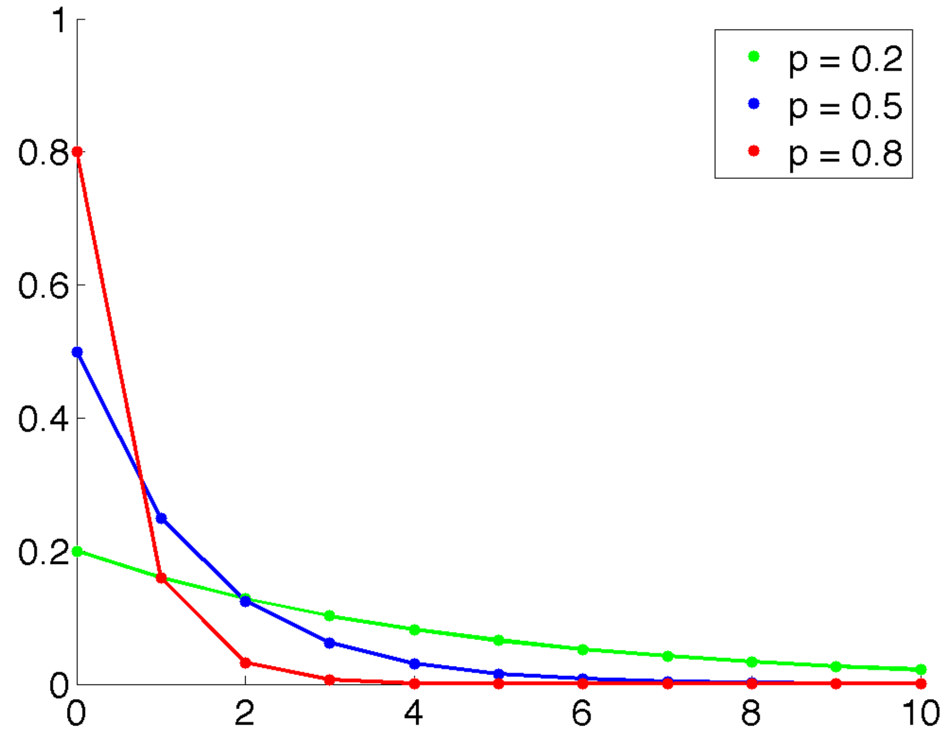
$$\mathrm{Geom}(x|p) = (1-p)^x p$$

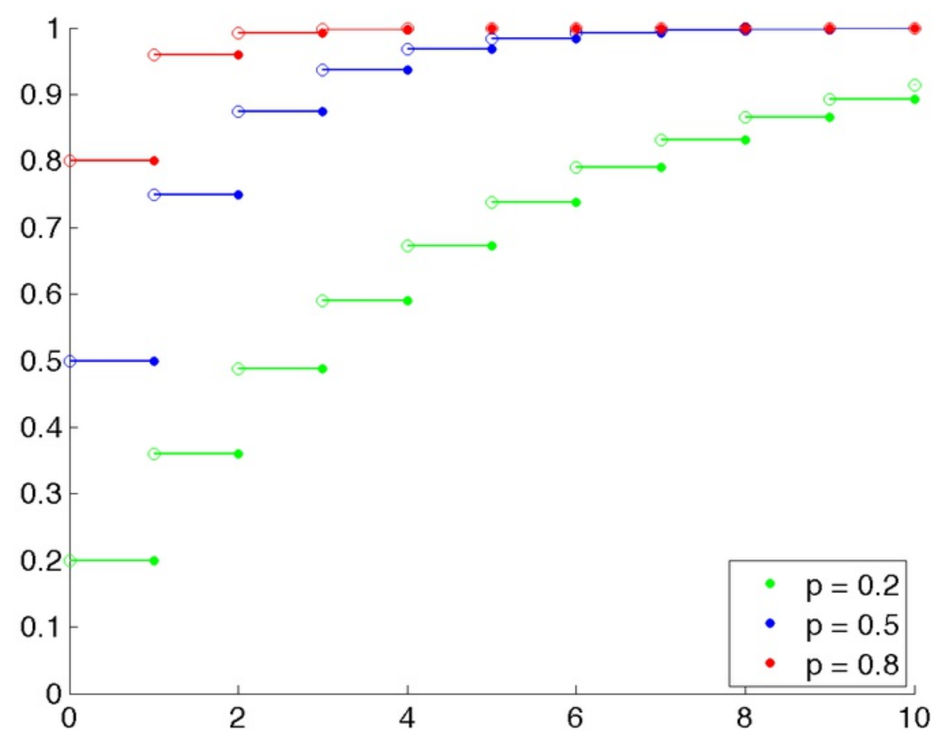
$$E[X] = 1/p - 1$$

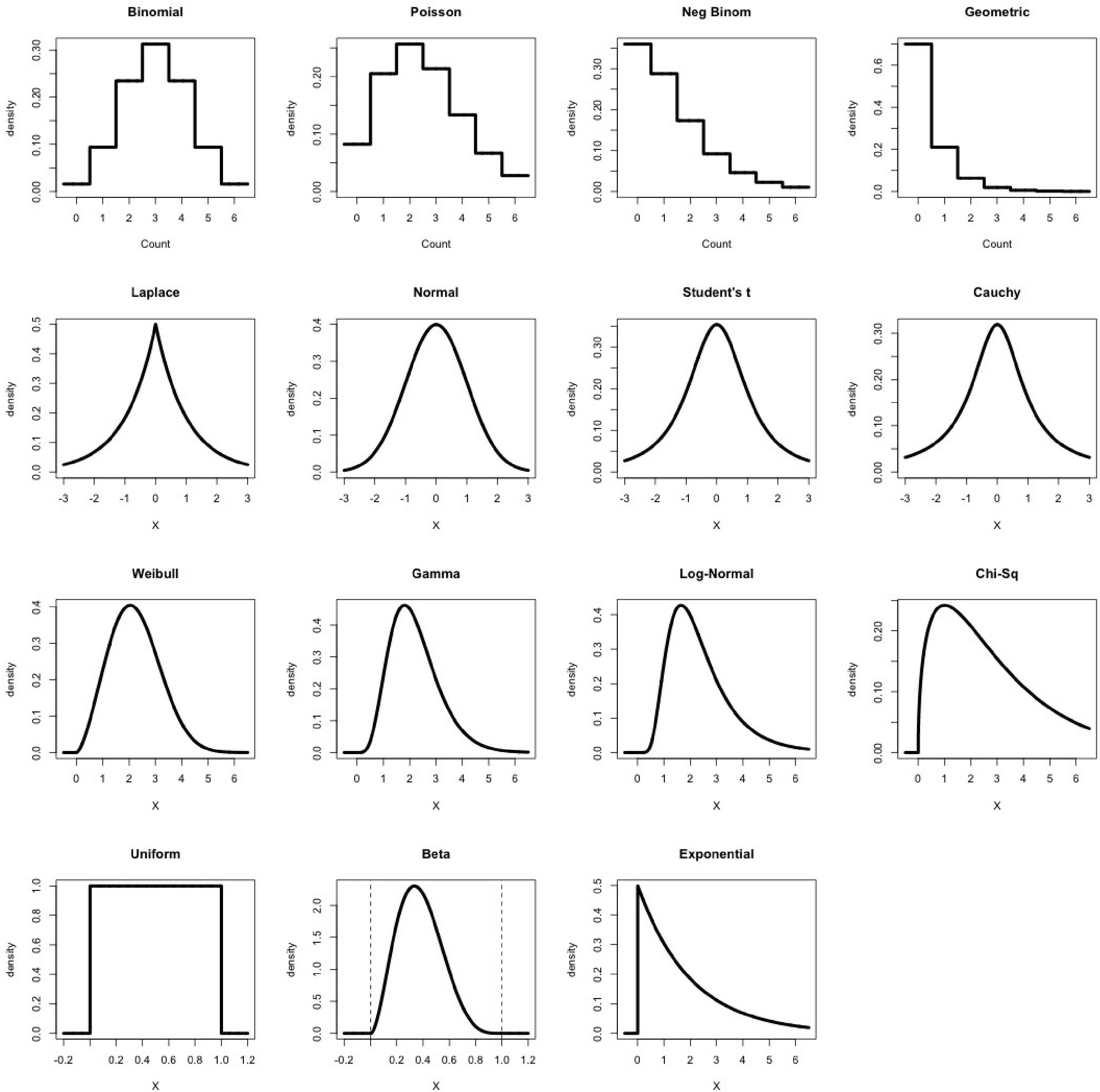
$$Var(X) = \frac{1-p}{p^2}$$

Number of trials required until success

Special case of the Negative Binomial
Geom(x|p) = NP(x|1,p)





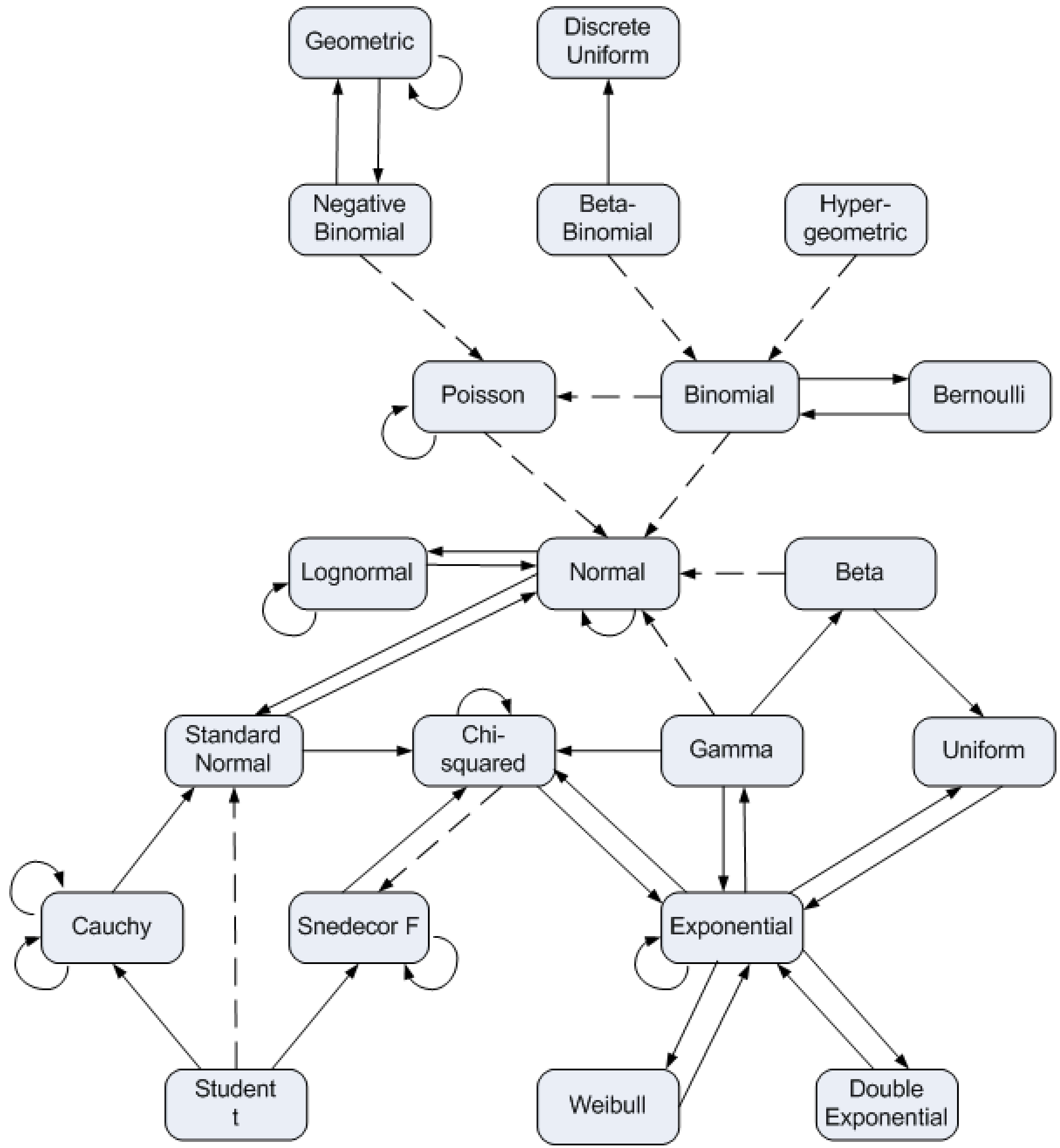


Binomial
density
Count
Poisson
density
Count
Neg Binom
density
Count
Geometric
density
Count
Laplace
density
X
Normal
density
X
Student's t
density
X
Cauchy
density
X
Weibull
density
X
Gamma
density
X
Log-Normal
density
X
Chi-Sq
density
X
Uniform
density
X
Beta
density
X
Exponential
density
X

TABLE 6.1. Common Probability Distributions and Their Characteristics

Name	*Equation*	*Bound*	*Interpretations and Relation to Other Distributions*
Discrete			
Bernoulli (p)	$p^x(1-p)^{1-x}$	[0,1]	Success ($x = 1$) with probability p. Binomial with $n = 1$.
Binomial (n,p)	$\binom{n}{x}p^x(1-p)^{n-x}$	[0 … N]	Number of successes, with probability p, given n trials.
Poisson (λ)	$\frac{\lambda^x}{x!}e^{-\lambda}$	0,∞	Number of events, occurring at rate λ, to occur over a fixed interval.
Negative binomial (n,p)	$\binom{x+n-1}{x}p^n(1-p)^x$	0,∞	Number of trials, with probability p, before n successes occur. Also a Poisson-Gamma mixture.
Geometric (p)	$p(1-p)^x$	0,∞	Number of trials needed before a success occurs. Special case of negative binomial with $n = 1$.
Continuous			
Uniform (a,b)	$\frac{1}{b-a}$	a,b	All values between a and b have equal probability.
Beta (α,β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	0,1	Probability of success given α out of $\alpha+\beta$ trials were successful. Beta (1,1) = Unif (0,1).
Exponential (λ)	$\lambda e^{-\lambda x}$	0,∞	The interval between events in a Poisson process at rate λ.
Laplace (μ,b)	$\frac{1}{2b}\exp\left(-\frac{\lvert x-\mu\rvert}{b}\right)$	−∞,∞	Two-sided exponential.
Weibull (λ,k)	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-\left(\frac{x}{\lambda}\right)^k}$	0,∞	Generalization of exponential, where λ changes according to k. Weibull (λ,1) = Exp (λ).
Gamma (a,r)	$\frac{r^a}{\Gamma(a)}x^{a-1}e^{-rx}$	0,∞	Sum of a Exp (r) variables, Gamma (1,r) = Exp (r). As a prior precision, Gamma (*sample size*/2,*sum of squares*/2). Gamma (n/2,2) = $\chi^2(n)$.
Normal (μ, σ)	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	−∞,∞	Maximum entropy distribution given a mean and variance.
Lognormal (μ,σ)	$\frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)$	0,∞	log(x) is Normally distributed. Note the mean is $e^{\mu+\sigma^2/2}$ not μ.
Student's t (n)	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}(1+x^{2/n})^{-\frac{n+1}{2}}$	−∞,∞	Normal with a Gamma distributed precision.
Cauchy (μ,γ)	$\frac{1}{\pi\gamma\left[1+\left(\frac{x-\mu}{\gamma}\right)^2\right]}$	−∞,∞	Ratio of two Normal variables. Has no mean and variance.
Chi-squared (n)	$\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}$	0,∞	Sum of squares of n standard Normal variables.

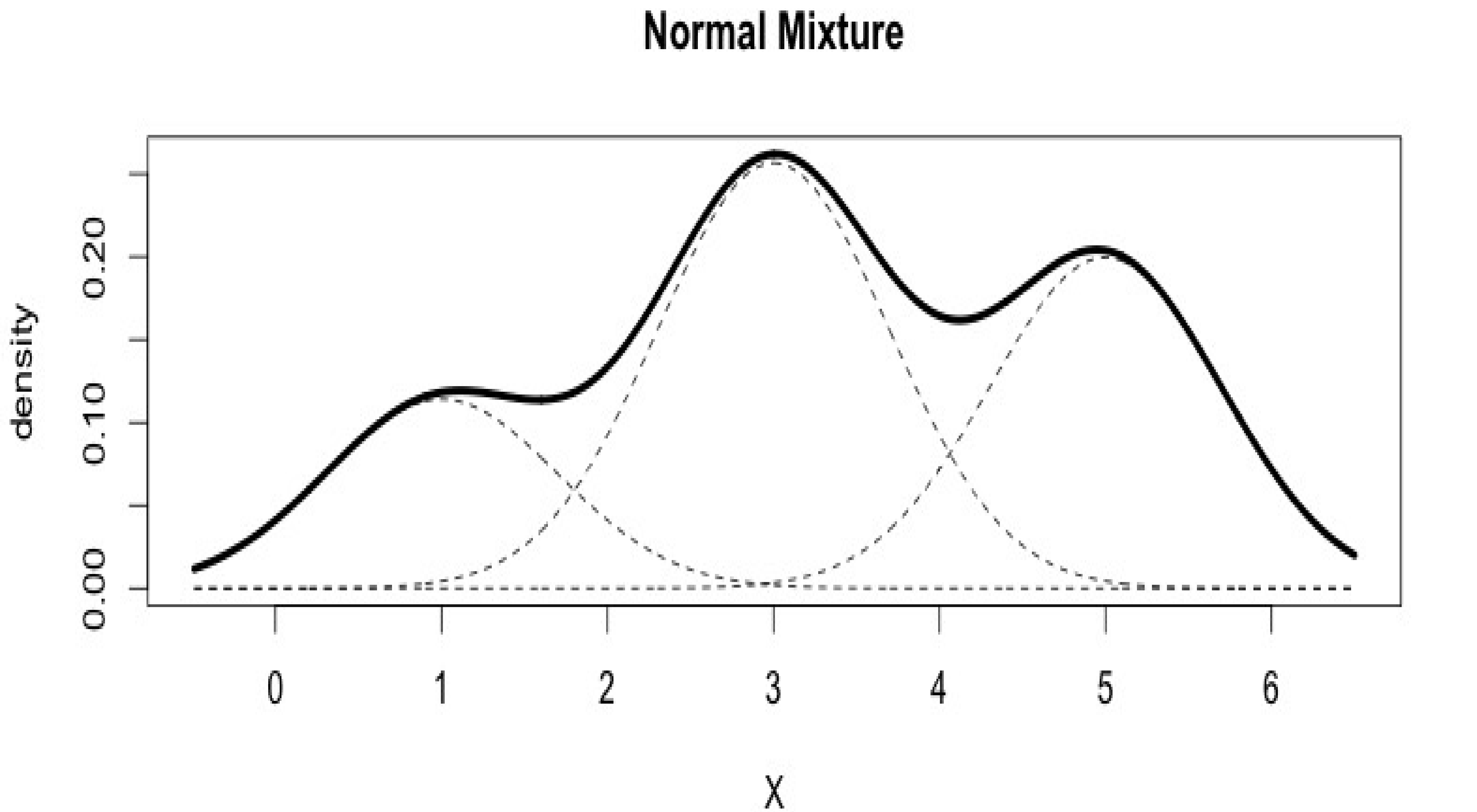
Note: All of the equations are parameterized with x as the random variable.


Geometric
Discrete Uniform
Negative Binomial
Beta-Binomial
Hyper-geometric
Poisson
Binomial
Bernoulli
Lognormal
Normal
Beta
Standard Normal
Chi-squared
Gamma
Uniform
Cauchy
Snedecor F
Exponential
Student t
Weibull
Double Exponential

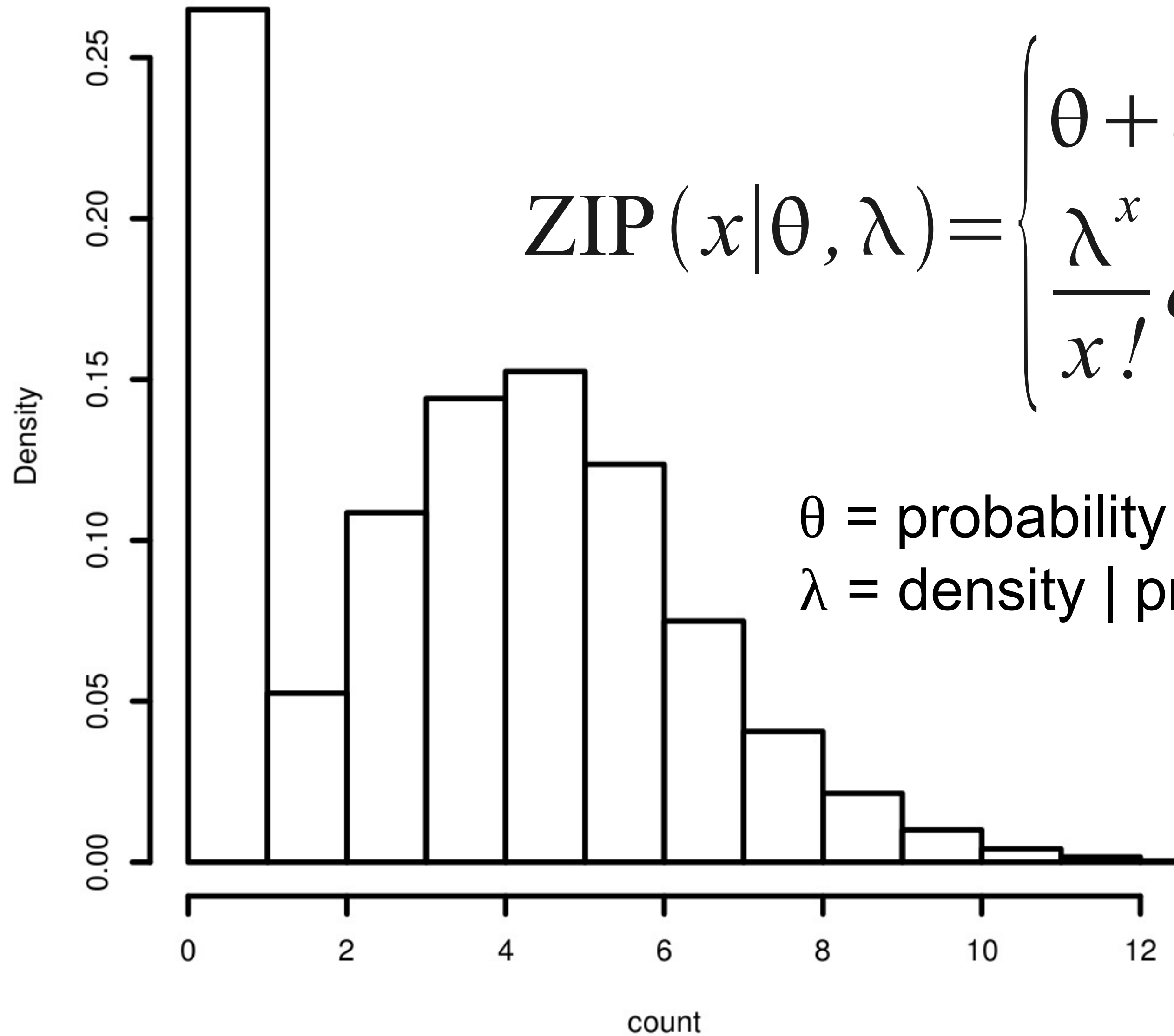
“Mixtures”

$$NB(x|s/r,s) = Pois(x|\mu)\,Gamma(\mu|s,r)$$

$$\mathrm{t}(x|s) = N(x|0,1/prec)\,Gamma(prec|s,s)$$


Normal Mixture
density
0.00
0.10
0.20
0
1
2
3
4
5
6
X

Zero Inflated Poisson



$$\mathrm{ZIP}(x|\theta,\lambda)=\begin{cases}\theta+e^{-\lambda} & x=0\\ \dfrac{\lambda^{x}}{x\,!}e^{-\lambda} & x>0\end{cases}$$

θ = probability of presence

λ = density | present

Looking forward...

$$P(\text{data} \mid \text{model})$$