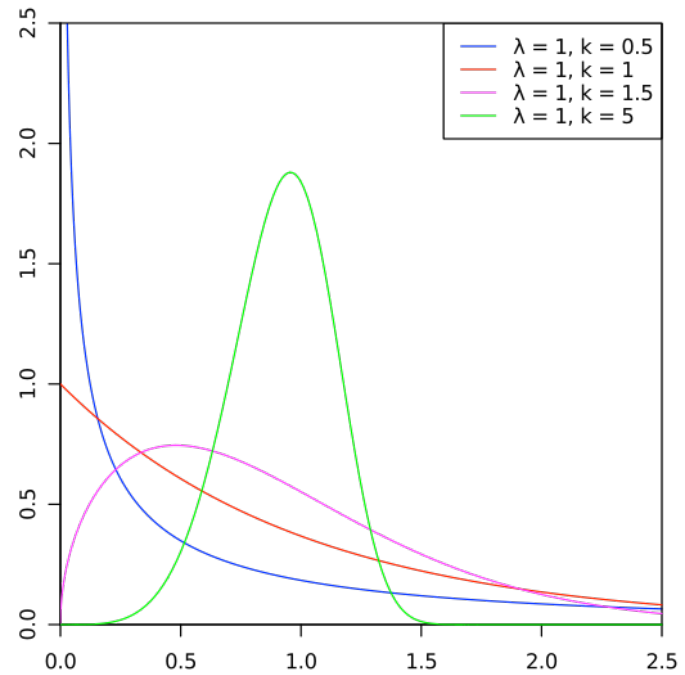
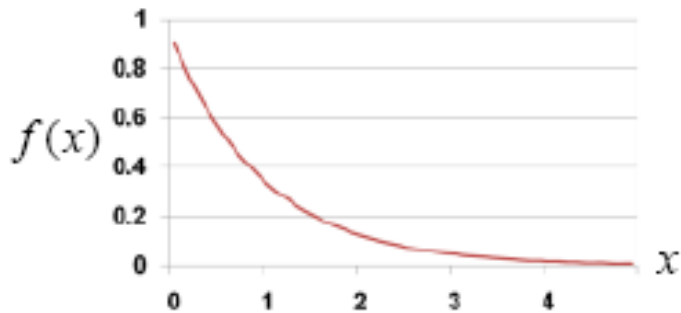


continuous random variables



Discrete random variable: takes values in a finite or countable set, e.g.

$X \in \{1, 2, \dots, 6\}$ with equal probability

X is positive integer i with probability 2^{-i}

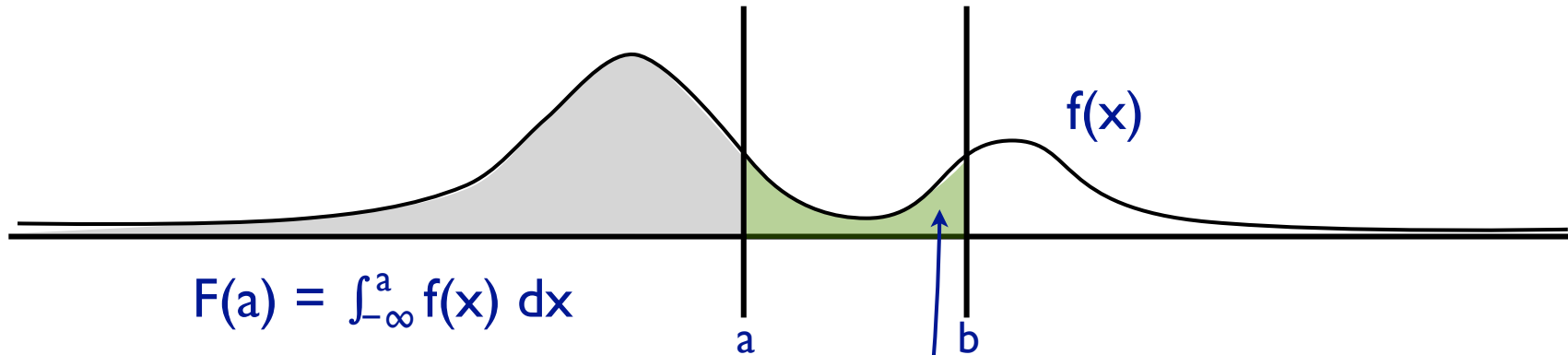
Continuous random variable: takes values in an uncountable set, e.g.

X is the weight of a random person (a real number)

X is a randomly selected point inside a unit square

X is the waiting time until the next packet arrives at the server

$f(x)$: the *probability density function* (or simply “density”)



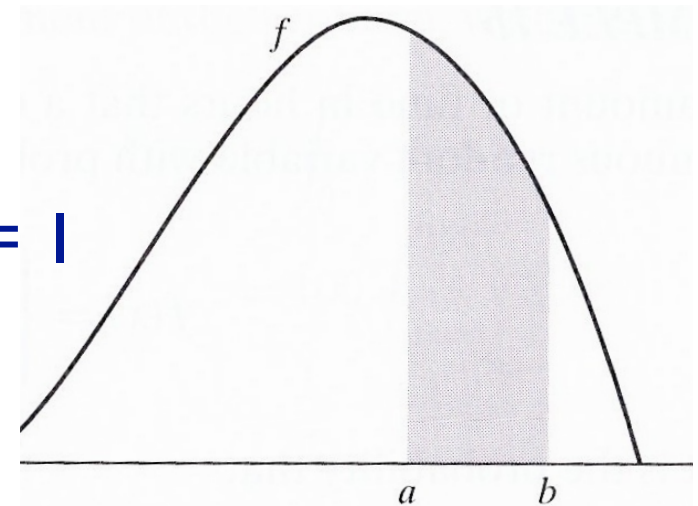
$P(X \leq a) = F(x)$: the *cumulative distribution function* (or simply “distribution”)

$P(a < X \leq b) = F(b) - F(a)$

Need $f(x) \geq 0$, & $\int_{-\infty}^{+\infty} f(x) dx (= F(+\infty)) = 1$

A key relationship:

$f(x) = \frac{d}{dx} F(x)$, since $F(a) = \int_{-\infty}^a f(x) dx$,



Densities are *not* probabilities

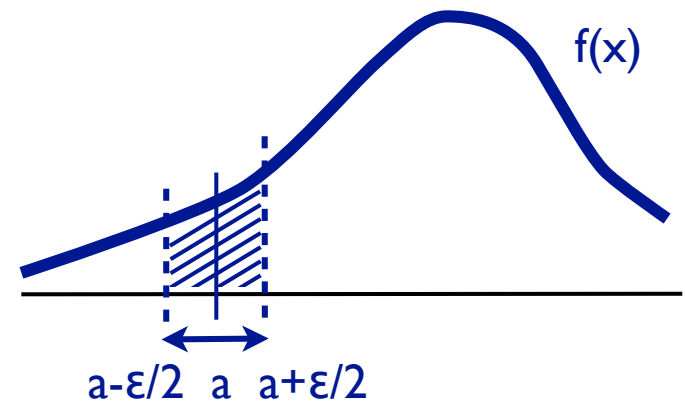
$$P(x = a) = P(a \leq X \leq a) = F(a) - F(a) = 0$$

I.e., the probability that a continuous random variable falls *at* a specified point is *zero*

$$P(a - \varepsilon/2 \leq X \leq a + \varepsilon/2) =$$

$$F(a + \varepsilon/2) - F(a - \varepsilon/2)$$

$$\approx \varepsilon \cdot f(a)$$



I.e., The probability that it falls *near* that point is proportional to the density; in a large random sample, expect more samples where density is higher (hence the name “density”).

Much of what we did with discrete r.v.s carries over almost unchanged, with $\sum_x \dots$ replaced by $\int \dots dx$

E.g.

For discrete r.v. X , $E[X] = \sum_x x p(x)$

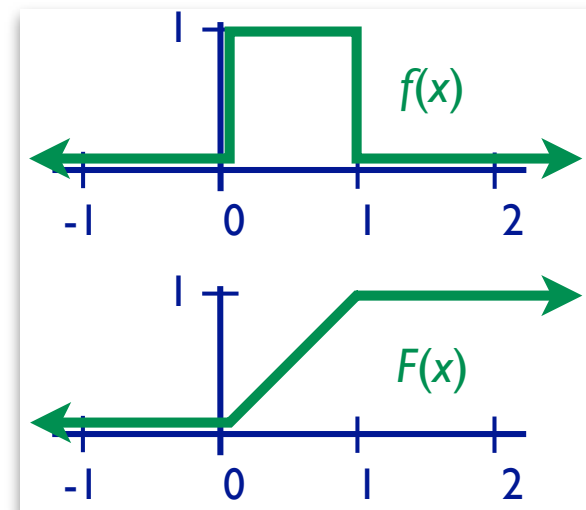
For continuous r.v. X , $E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$

Why?

(a) We define it that way

(b) The probability that X falls “near” x , say within $x \pm dx/2$, is $\approx f(x)dx$, so the “average” X should be $\approx \sum x f(x)dx$ (summed over grid points spaced dx apart on the real line) and the limit of that as $dx \rightarrow 0$ is $\int x f(x)dx$

$$\text{Let } f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$



$$F(a) = \int_{-\infty}^a f(x) dx$$

$$= \begin{cases} 0 & \text{if } a \leq 0 \\ a & \text{if } 0 < a \leq 1 \text{ (since } a = \int_0^a 1 dx \text{)} \\ 1 & \text{if } 1 < a \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 dx = \left. \frac{x^3}{3} \right|_0^1 = \frac{1}{3}$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \quad (\sigma \approx 0.29)$$

Linearity

$$E[aX+b] = aE[X]+b$$

$$E[X+Y] = E[X]+E[Y]$$

still true, just as
for discrete

Functions of a random variable

$$E[g(X)] = \int g(x)f(x)dx$$

just as for discrete,
but w/integral

Definition is same as in the discrete case

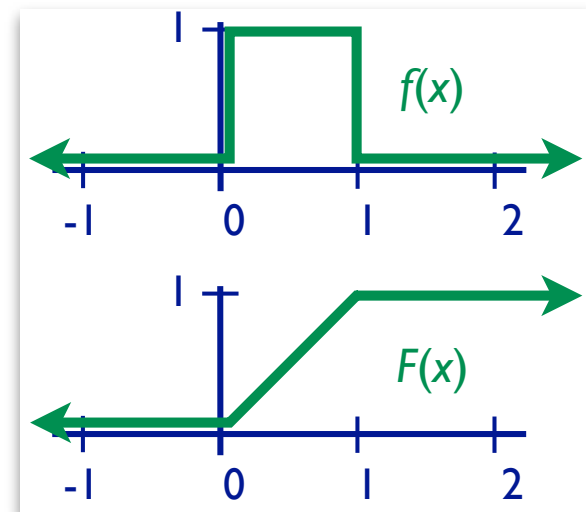
$$\text{Var}[X] = E[(X-\mu)^2] \quad \text{where } \mu = E[X]$$

Identity still holds:

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

proof “same”

$$\text{Let } f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$



$$F(a) = \int_{-\infty}^a f(x) dx$$

$$= \begin{cases} 0 & \text{if } a \leq 0 \\ a & \text{if } 0 < a \leq 1 \text{ (since } a = \int_0^a 1 dx \text{)} \\ 1 & \text{if } 1 < a \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 dx = \left. \frac{x^3}{3} \right|_0^1 = \frac{1}{3}$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \quad (\sigma \approx 0.29)$$

continuous random variables: summary

Continuous random variable X has density $f(x)$, and

$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx$$

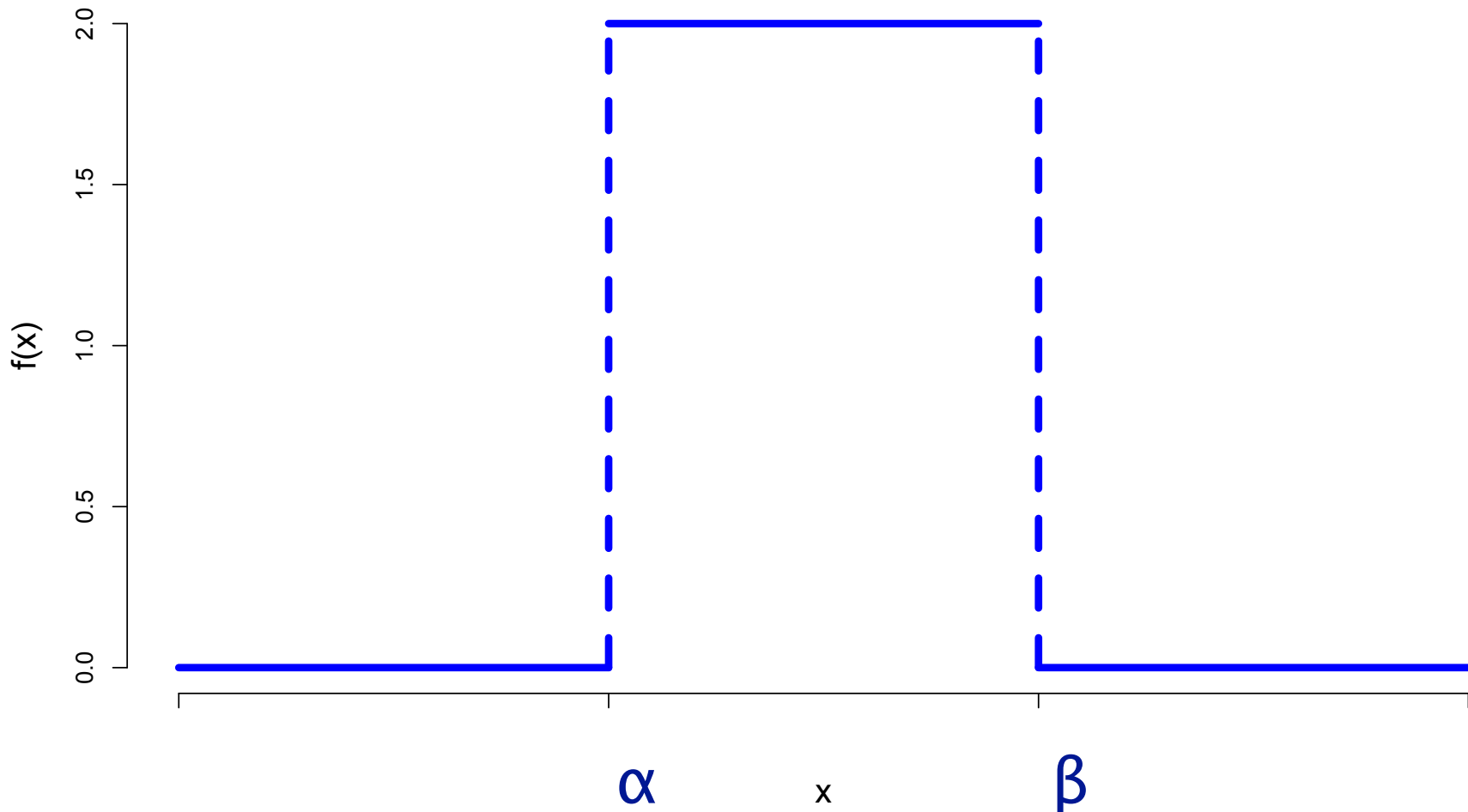
$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

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

uniform random variable

$$X \sim \text{Uni}(\alpha, \beta) \text{ is uniform in } [\alpha, \beta] \quad f(x) = \begin{cases} \frac{1}{\beta - \alpha} & x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

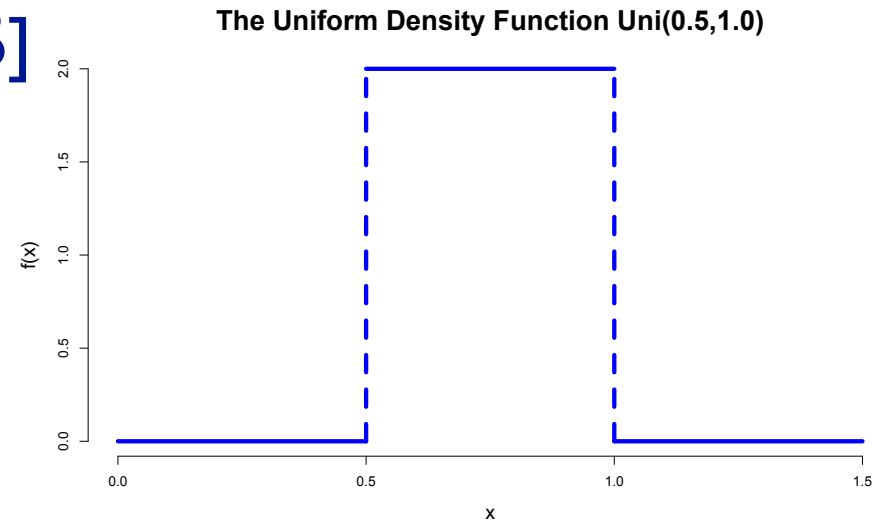
The Uniform Density Function Uni(0.5,1.0)



uniform random variable

$X \sim \text{Uni}(\alpha, \beta)$ is uniform in $[\alpha, \beta]$

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



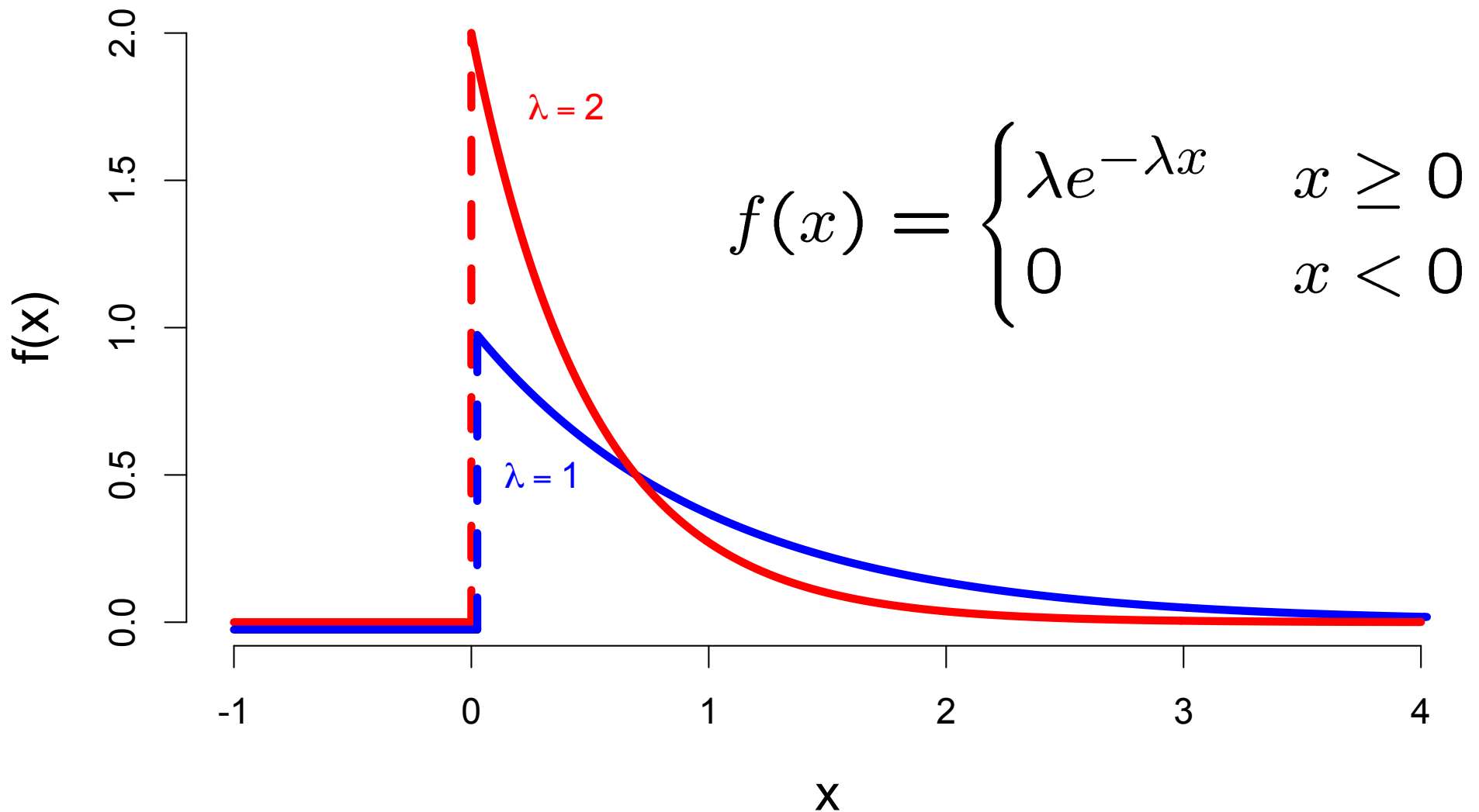
$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx = \frac{b - a}{\beta - \alpha}$$

if $\alpha \leq a \leq b \leq \beta$:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx = \frac{\alpha + \beta}{2}$$

$X \sim \text{Exp}(\lambda)$

The Exponential Density Function



$X \sim \text{Exp}(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

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

$$\Pr(X \geq t) = e^{-\lambda t} = 1 - F(t)$$

Memorylessness:

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

Radioactive decay: How long until the next alpha particle?

Customers: how long until the next customer/packet arrives at the checkout stand/server?

Buses: How long until the next #71 bus arrives on the Ave?

Yes, they have a schedule, but given the vagaries of traffic, riders with-bikes-and-baby-carriages, etc., can they stick to it?

Relation to the Poisson:

Poisson: *how many* events in a *fixed time*;

Exponential: *how long* until the *next event*

Relation to geometric: Geometric is discrete analog:

How long to a Head, 1 flip per sec, prob p vs

How long to a Head, 2 flips per sec, prob $p/2$, ...

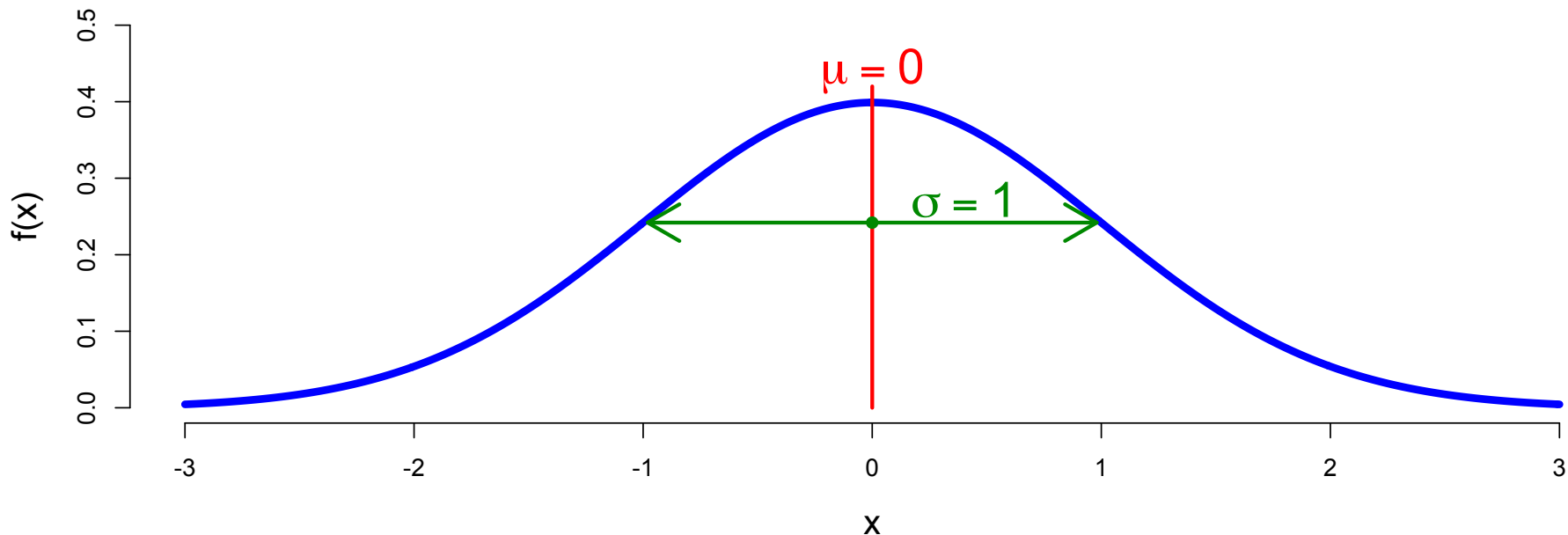
Limit is exponential with parameter p

X is a normal (aka Gaussian) random variable $X \sim N(\mu, \sigma^2)$

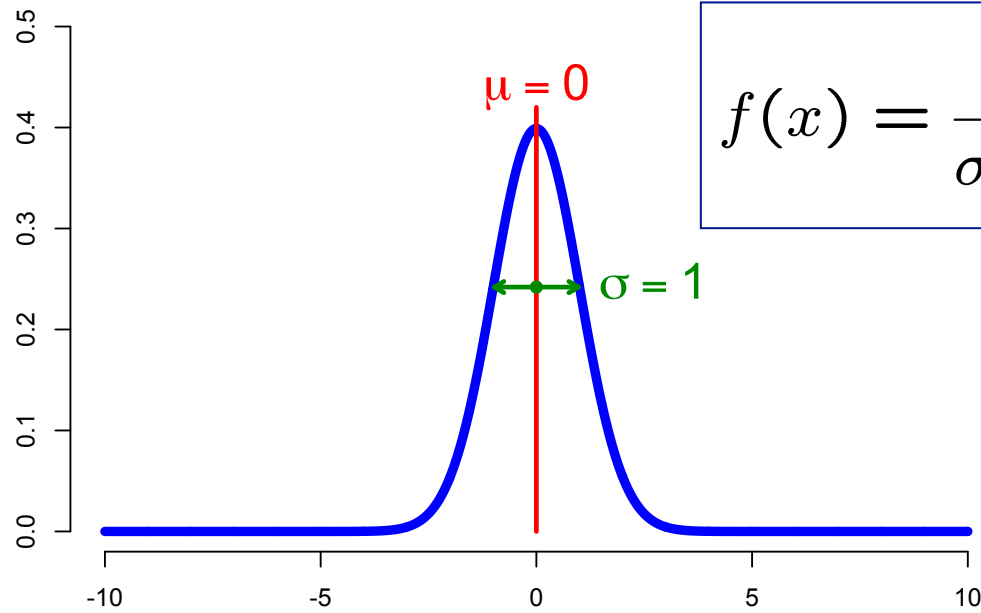
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$E[X] = \mu \quad \text{Var}[X] = \sigma^2$$

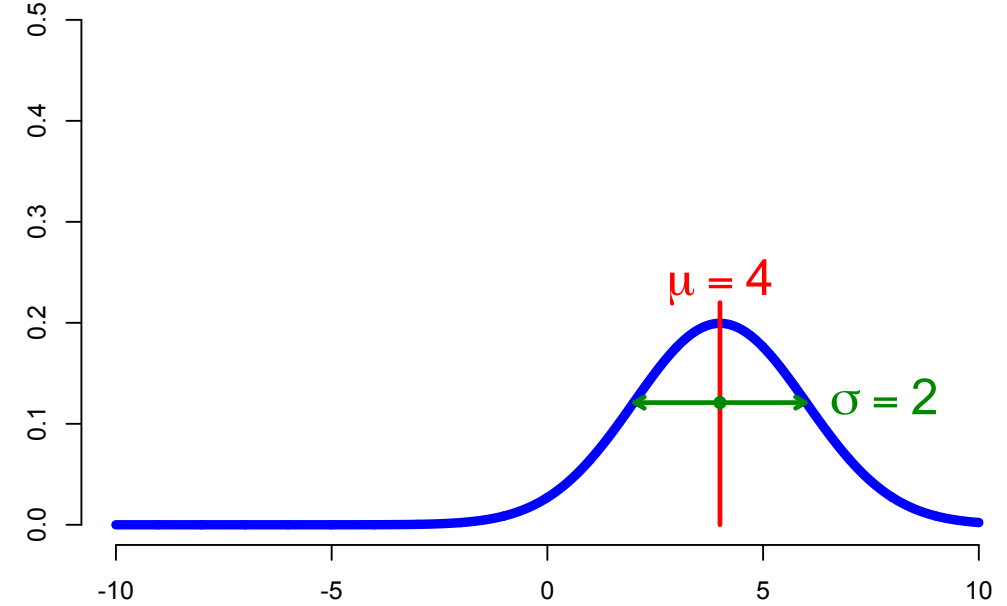
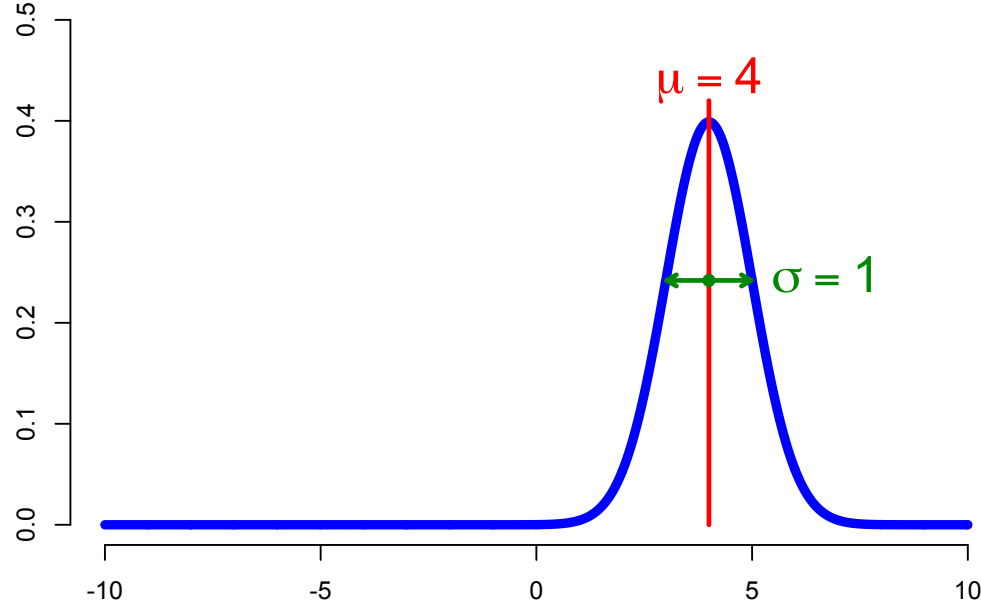
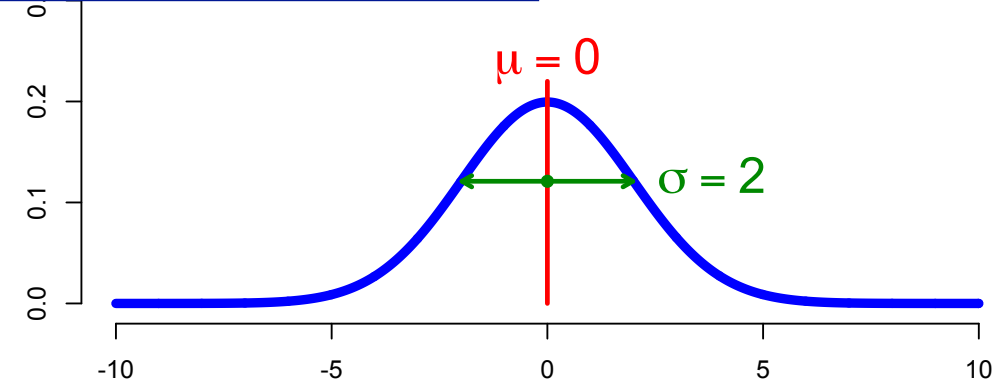
The Standard Normal Density Function



changing μ, σ



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



density at μ is $\approx .399/\sigma$

X is a normal random variable $X \sim N(\mu, \sigma^2)$

$$Y = aX + b$$

$$E[Y] = E[aX + b] = a\mu + b$$

$$\text{Var}[Y] = \text{Var}[aX + b] = a^2\sigma^2$$

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

Important special case: $Z = (X - \mu) / \sigma \sim N(0, 1)$

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

$Z \sim N(0, 1)$ “*standard (or unit) normal*”

Use $\Phi(z)$ to denote CDF, i.e.

$$\Phi(z) = \Pr(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

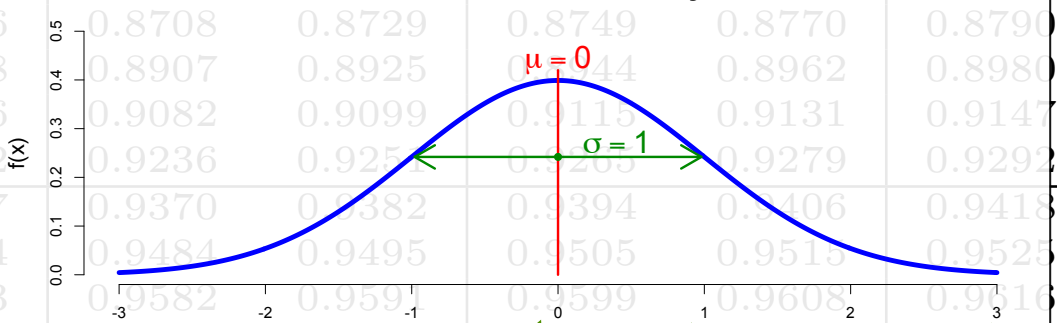
no closed form ☹

Table of the Standard Normal Cumulative Distribution Function $\Phi(Z)$

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7122 | 0.7157 | 0.7190 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8529 | 0.8549 | 0.8568 | 0.8587 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9685 | 0.9692 | 0.9699 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 |

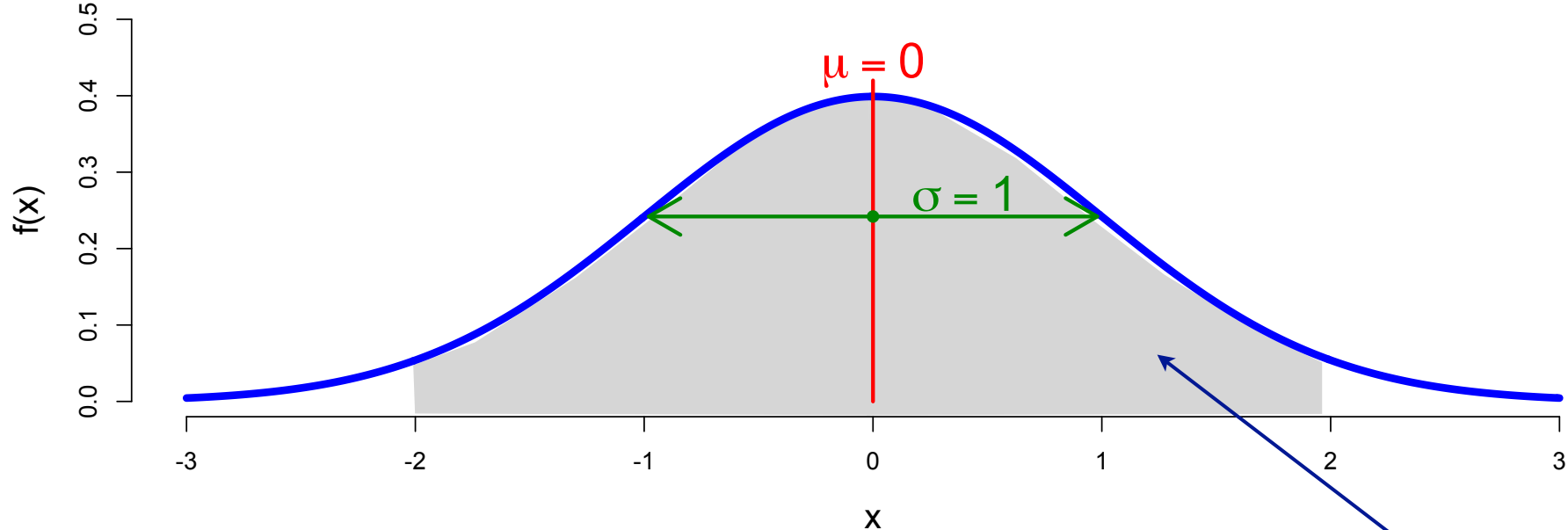
$\Phi(.46)$

The Standard Normal Density Function



E.g., see B&T p155, p531

The Standard Normal Density Function



If $Z \sim N(\mu, \sigma)$ what is $P(\mu - \sigma < Z < \mu + \sigma)$?

$$P(\mu - \sigma < Z < \mu + \sigma) = \Phi(1) - \Phi(-1) \approx 68\%$$

$$P(\mu - 2\sigma < Z < \mu + 2\sigma) = \Phi(2) - \Phi(-2) \approx 95\%$$

$$P(\mu - 3\sigma < Z < \mu + 3\sigma) = \Phi(3) - \Phi(-3) \approx 99\%$$

normal approximation to binomial

$X \sim \text{Bin}(n,p)$

$$E[X] = np \quad \text{Var}[X] = np(1-p)$$

Poisson approx: good for n large, p small (np constant)

Normal approx: For large n , (p stays fixed):

$$X \approx Y \sim N(E[X], \text{Var}[X]) = N(np, np(1-p))$$

Normal approximation good when $np(1-p) \geq 10$

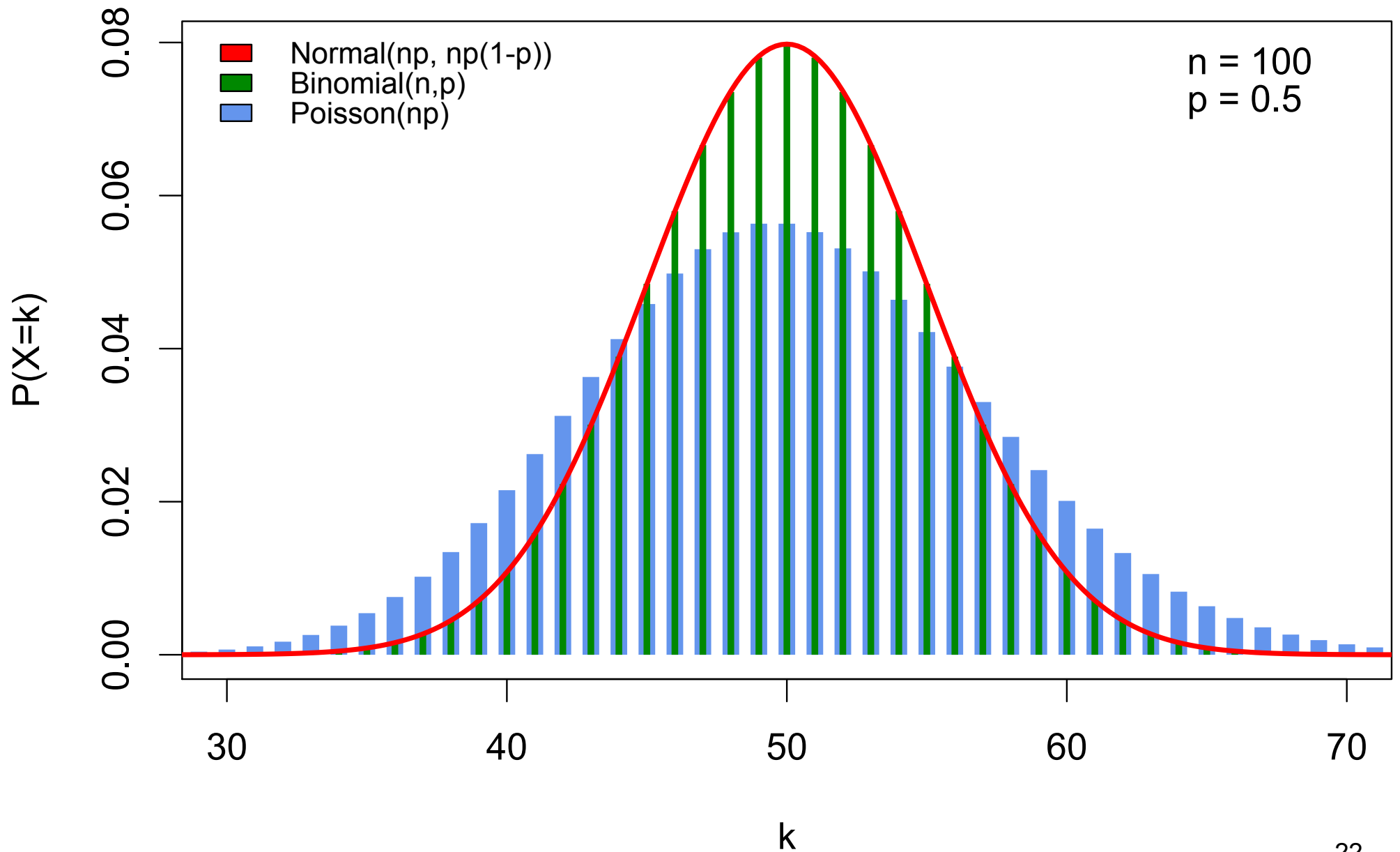
DeMoivre-Laplace Theorem:

Let S_n = number of successes in n trials (with prob. p).

Then, as $n \rightarrow \infty$:

$$Pr \left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) \longrightarrow \Phi(b) - \Phi(a)$$

normal approximation to binomial



normal approximation to binomial

Fair coin flipped 40 times. Probability of 20 heads?

Exact answer:

$$P(X = 20) = \binom{40}{20} \left(\frac{1}{2}\right)^{40} \approx \boxed{0.1254}$$

Normal approximation:

$$\begin{aligned} P(X = 20) &= P(19.5 \leq X < 20.5) \\ &= P\left(\frac{19.5 - 20}{\sqrt{10}} \leq \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right) \\ &\approx P\left(-0.16 \leq \frac{X - 20}{\sqrt{10}} < 0.16\right) \\ &\approx \Phi(0.16) - \Phi(-0.16) \approx \boxed{0.1272} \end{aligned}$$

the central limit theorem (CLT)

Consider i.i.d. (independent, identically distributed) random vars X_1, X_2, X_3, \dots

X_i has $\mu = E[X_i]$ and $\sigma^2 = \text{Var}[X_i]$

As $n \rightarrow \infty$,

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0, 1)$$

Restated: As $n \rightarrow \infty$,

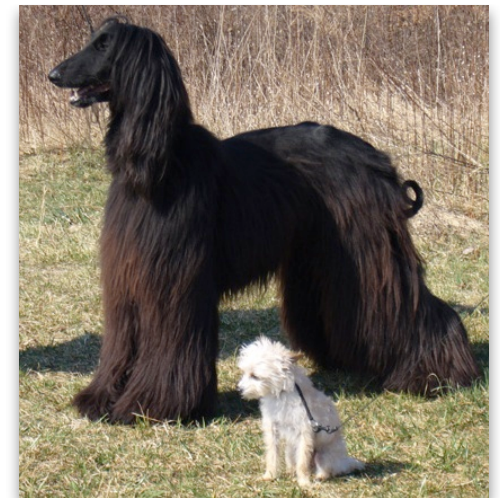
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

How tall are you? Why?



Credit: Annie Leibovitz, © 1987 ?

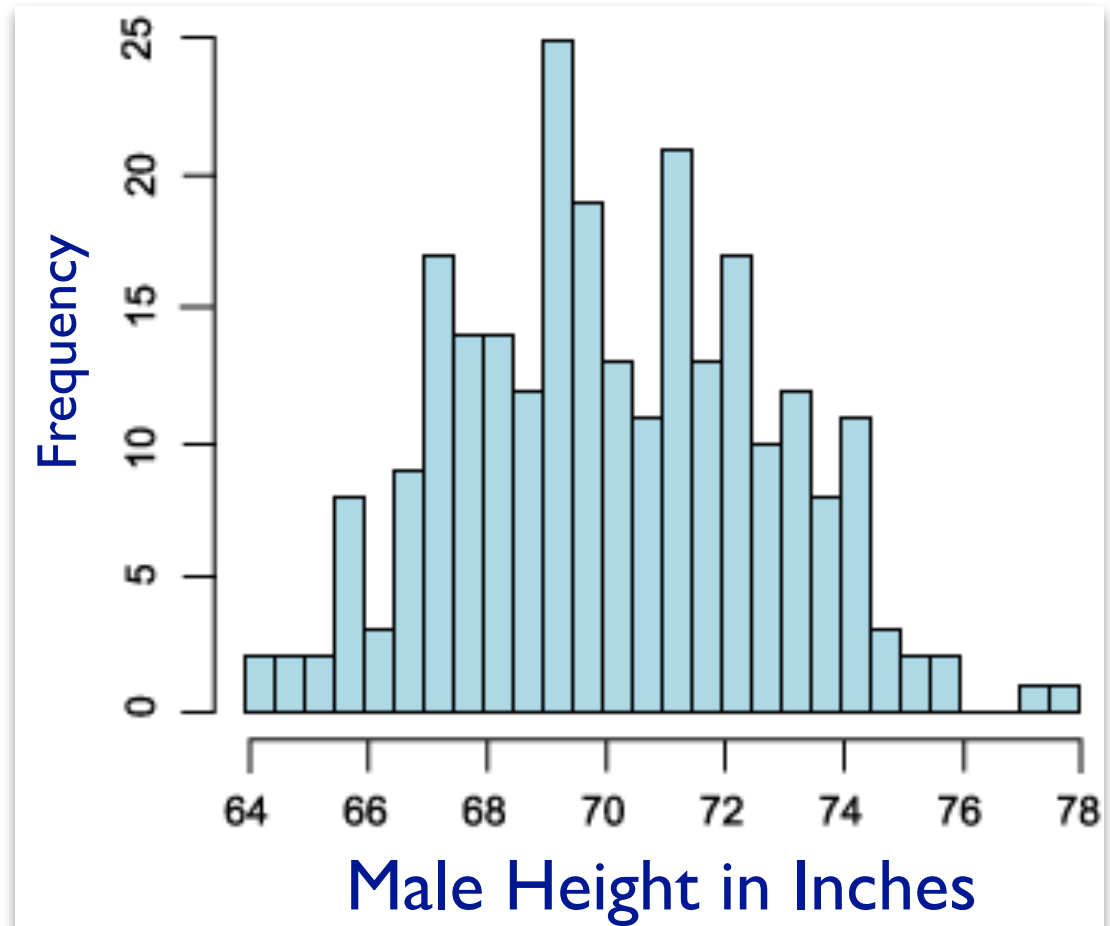
Willie Shoemaker & Wilt Chamberlain



Human height is approximately normal.

Why might that be true?

R.A. Fisher (1918) noted it would follow from CLT if height were the sum of many independent random effects, e.g. many genetic factors (plus some environmental ones like diet). *I.e., suggested part of mechanism by looking at shape of the curve. (WAY before anyone really knew what genes were...)*



Meta-analysis of Dense Genecentric Association Studies Reveals Common and Uncommon Variants Associated with Height, Lanktree, et al.

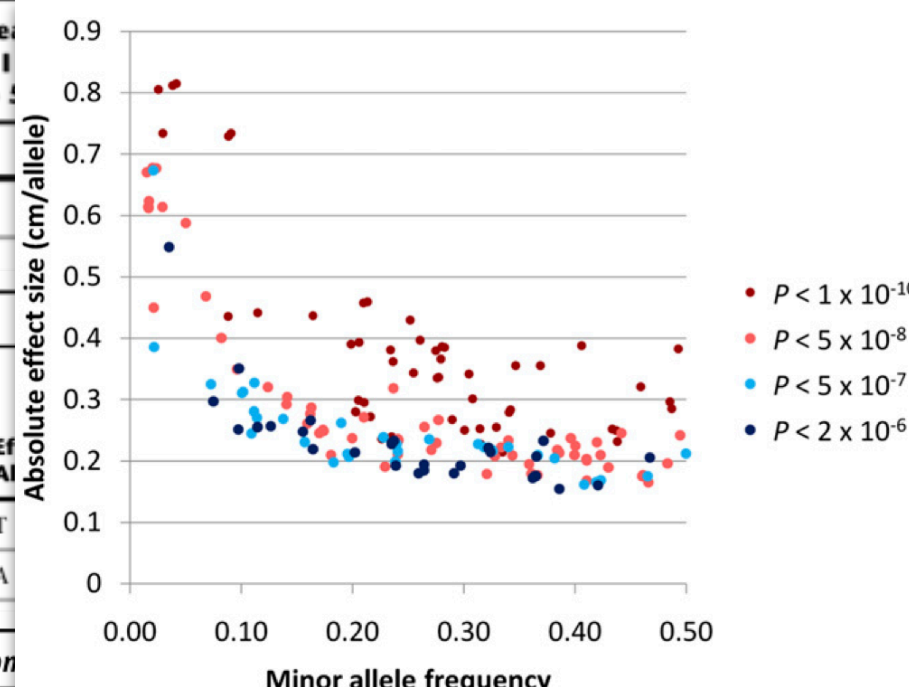
The American Journal of Human Genetics 88, 6–18, January 7, 2011

Table 1. Sixty-Four Loci Showing Significant Evidence for Association with Adult Height, Identified with the Use of the IBC Array

| Locus Rank | Chr. | Candidate Gene ^a | SNP ^a | Effect Allele | MAF | European Ancestry Phase I (up to 53,394) | | |
|------------|-------|-----------------------------|------------------|---------------|------|--|---|----------------|
| | | | | | | Effect | p | I ² |
| 1 | 7q22 | CDK6 | rs4272 | A | 0.21 | -0.46 | | |
| 2 | 6p21 | HMGA1 | rs1150781 | C | 0.09 | 0.72 | | |
| 3 | 12q15 | HMGA2 | | | | | | |
| 4 | 20q11 | MMP24 | | | | | | |
| 5 | 17q23 | MAP3K3 | | | | | | |
| 6 | 17q24 | GH1-GH2 | | | | | | |
| 7 | 1p36 | MEAP2 | | | | | | |
| 8 | 15q26 | IGF1R | | | | | | |
| 9 | 7p22 | GNA12 | | | | | | |
| 10 | 17q23 | TBX2 | | | | | | |
| 11 | 12q22 | SOCS2 | | | | | | |
| 12 | 9q22 | PTCH1 | | | | | | |
| 13 | 14q11 | NFATC4 | | | | | | |
| 14 | 15q26 | ACAN | | | | | | |
| 15 | 2q24 | NPPC | | | | | | |
| 16 | 6p21 | PPARD | | | | | | |
| 17 | 20q11 | MYH7B | | | | | | |
| 18 | 19q13 | IL11 | | | | | | |

| Table 1. Continued | | | | | | | | | |
|--------------------|-------|-----------------------------|------------------|---------------|------|--------|------------------------|----------------|--|
| Locus Rank | Chr. | Candidate Gene ^a | SNP ^a | Effect Allele | MAF | Effect | p | I ² | European Ancestry Phase I (up to 53,394) |
| 28 | 2p23 | GCKR | rs780094 | T | 0.10 | 0.23 | 3.8 × 10 ⁻⁵ | 2 | |
| 29 | 1q41 | TGFB2 | rs900 | A | 0.10 | -0.25 | 2.7 × 10 ⁻⁴ | 0 | |
| 30 | 20q11 | CDK5RAP1 | | | | | | | |
| 31 | 2p12 | EIF2AK3 | | | | | | | |
| 32 | 19p13 | INSR | | | | | | | |
| 33 | 6q25 | ESR1 | | | | | | | |
| 34 | 2q37 | DIS3L2 | | | | | | | |
| 35 | 2q35 | PLCD4 | | | | | | | |
| 36 | 1p36 | RPS6KA1 | | | | | | | |
| 37 | 15q21 | CYP19A1 | | | | | | | |
| 38 | 5q31 | SLC22A5 | | | | | | | |
| 39 | 7p15 | JAZF1 | | | | | | | |
| 40 | 17p13 | POLR2A | | | | | | | |
| 41 | 1p22 | PKN2 | | | | | | | |
| 42 | 7q22 | CNOT4 | | | | | | | |

| Table 1. Continued | | | | | | | | | |
|--------------------|-------|-----------------------------|------------------|---------------|------|--------|------------------------|----------------|--|
| Locus Rank | Chr. | Candidate Gene ^a | SNP ^a | Effect Allele | MAF | Effect | p | I ² | European Ancestry Phase I (up to 53,394) |
| 54 | 1p22 | COL24A1 | rs2046159 | A | 0.16 | 0.23 | 3.8 × 10 ⁻⁵ | 2 | |
| 55 | 1q23 | DUSP23 | rs1129923 | A | 0.10 | -0.25 | 2.7 × 10 ⁻⁴ | 0 | |
| 56 | 10q22 | MAT1A | rs7087728 | A | 0.18 | 0.22 | 2.2 × 10 ⁻⁴ | 0 | |
| 57 | 2p15 | PPP3R1 | rs1822469 | T | 0.41 | -0.14 | 7.8 × 10 ⁻⁴ | 9 | |
| 58 | 7q36 | ATG9B | rs1800783 | A | 0.38 | -0.16 | 2.0 × 10 ⁻⁴ | 0 | |
| 59 | 14q11 | BCL2L2 | rs3210043 | A | 0.16 | 0.25 | 9.7 × 10 ⁻⁶ | 0 | |
| 60 | 4p14 | RFC1 | rs11096991 | T | 0.35 | 0.15 | 3.6 × 10 ⁻⁴ | 0 | |
| 61 | 6p21 | HMGA1 | rs1150781 | C | 0.09 | 0.72 | 1.5 × 10 ⁻³ | 1 | |



(and hundreds more probably exist)

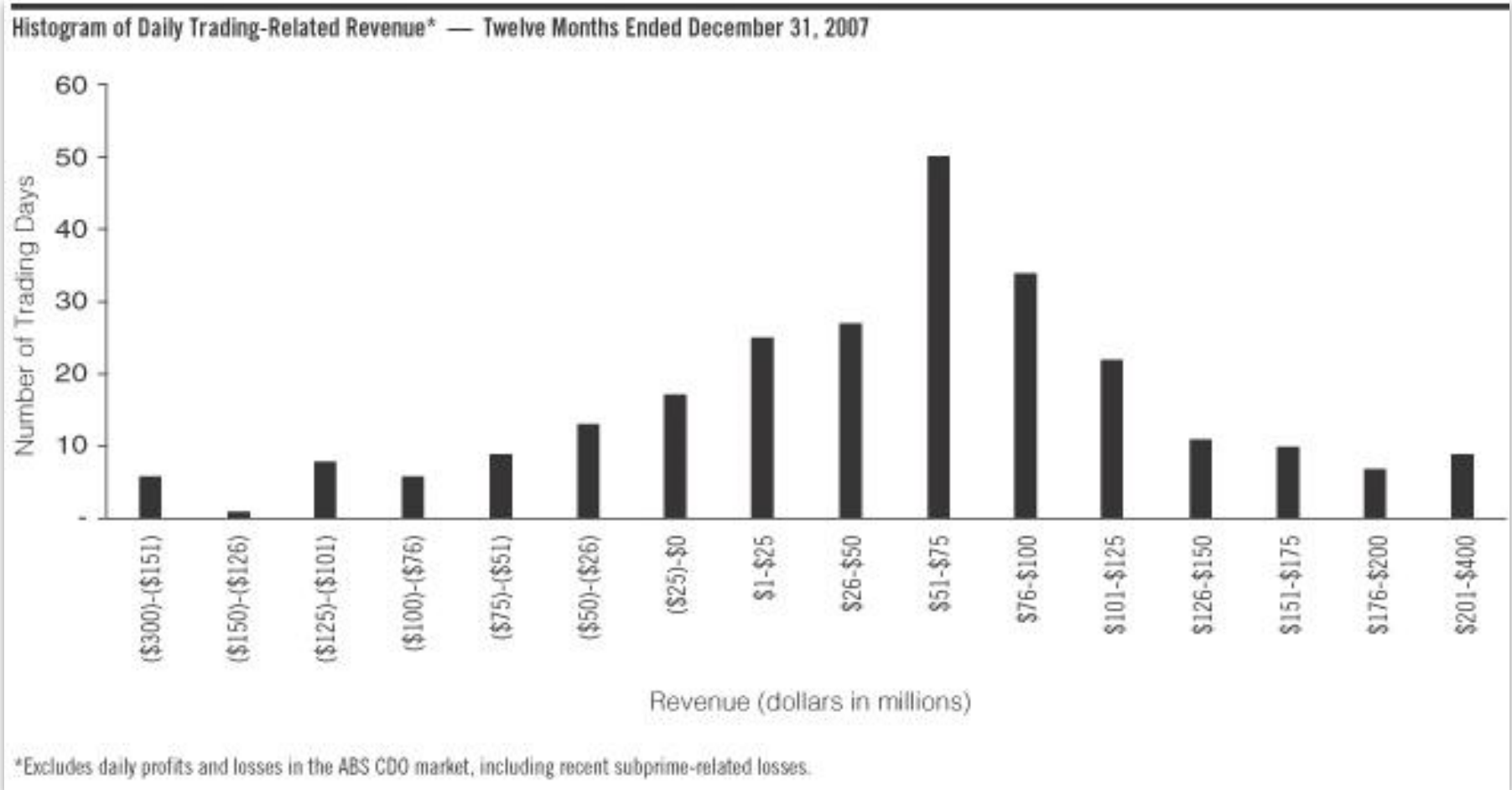
Table 1. Continued

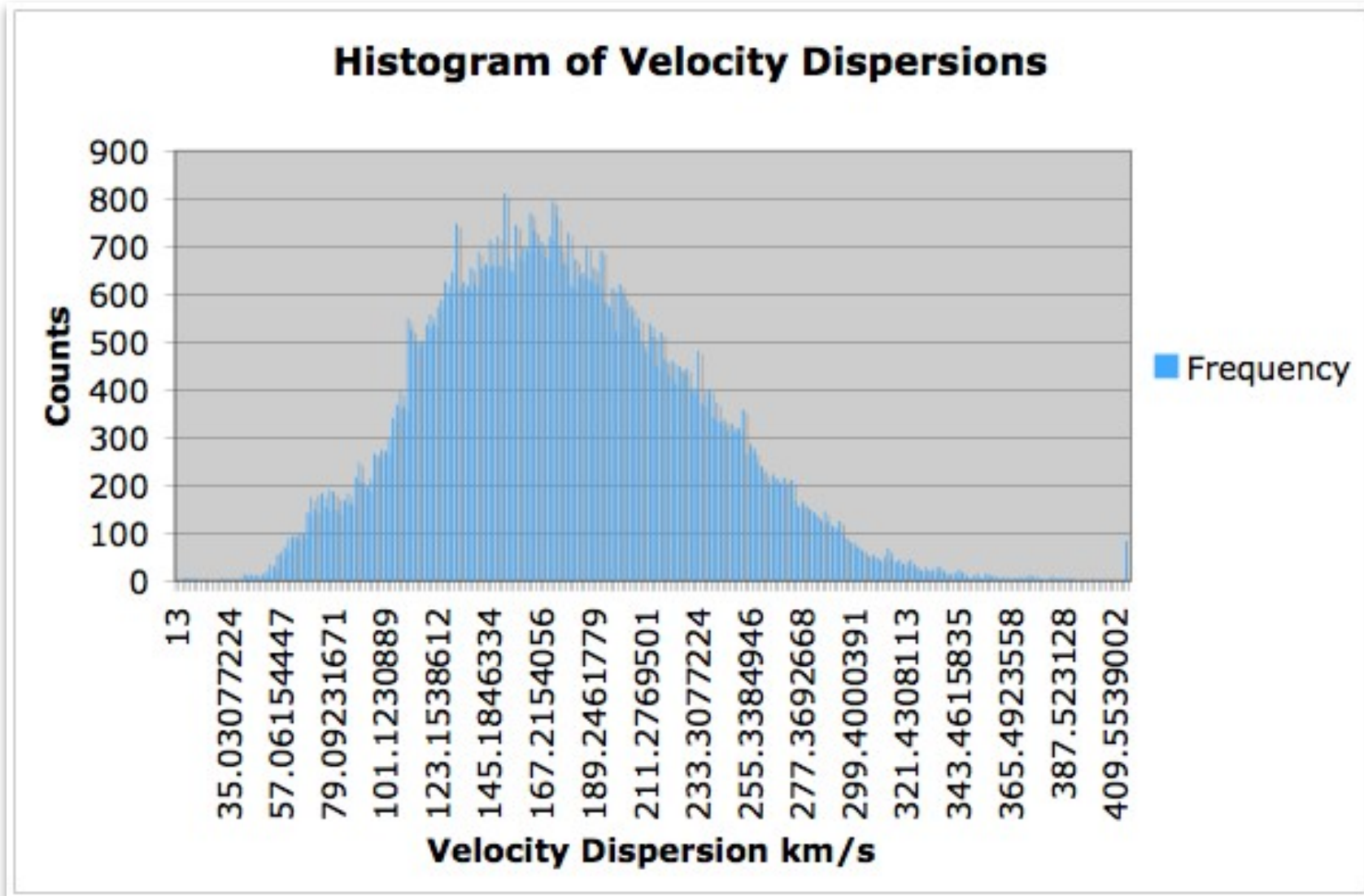
Table 1. Continued

in the real world...



in the real world...





pdf and cdf

$$f(x) = \frac{d}{dx} F(x) \quad F(a) = \int_{-\infty}^a f(x) dx$$

sums become integrals, e.g.

$$E[X] = \sum_x x p(x) \quad E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

most familiar properties still hold, e.g.

$$E[aX+bY+c] = aE[X]+bE[Y]+c$$

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

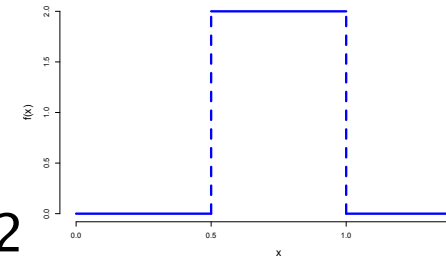
Three important examples

$X \sim \text{Uni}(\alpha, \beta)$ uniform in $[\alpha, \beta]$

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

$$E[X] = (\alpha + \beta) / 2$$

$$\text{Var}[X] = (\beta - \alpha)^2 / 12$$

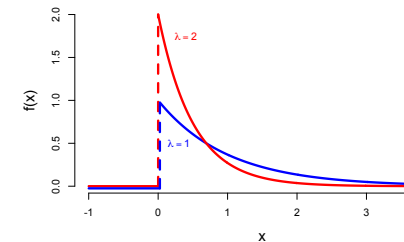


$X \sim \text{Exp}(\lambda)$ exponential

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$E[X] = \frac{1}{\lambda}$$

$$\text{Var}[X] = \frac{1}{\lambda^2}$$



$X \sim \text{N}(\mu, \sigma^2)$ normal (aka Gaussian)

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

$$E[X] = \mu$$

$$\text{Var}[X] = \sigma^2$$

