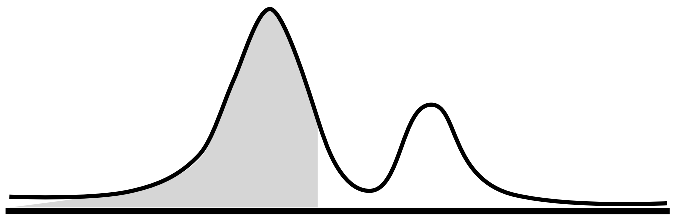
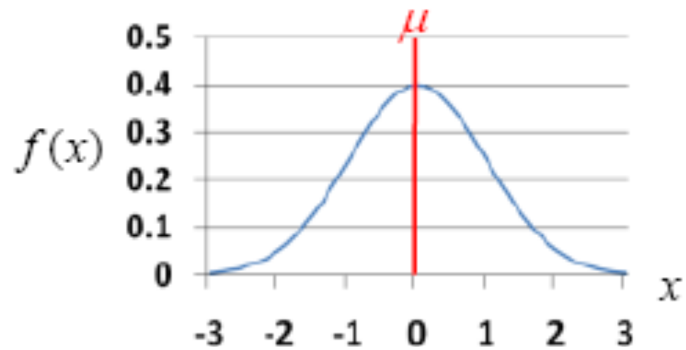
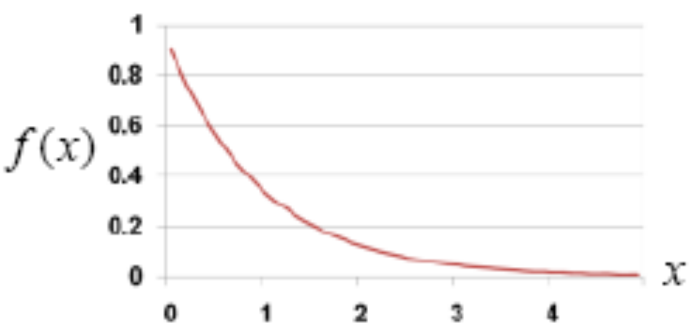
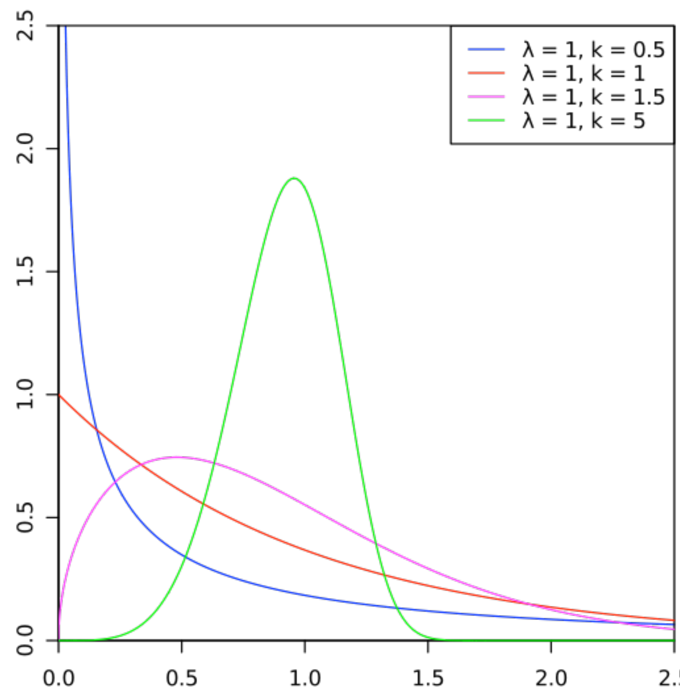




7. continuous random variables





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

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

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

probability mass function assigns probabilities to points

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

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

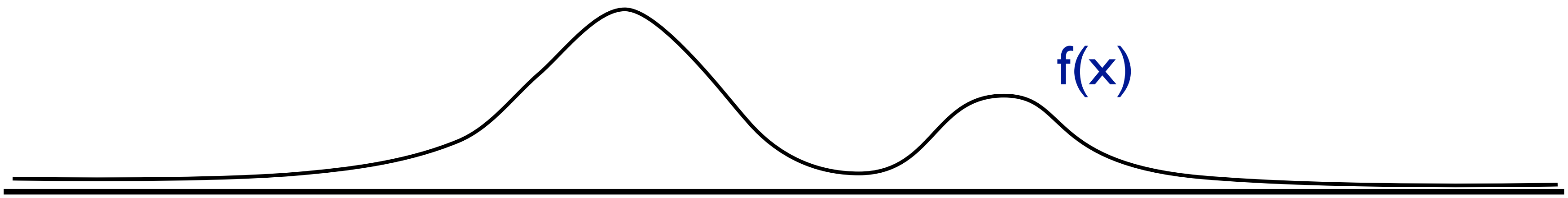
X is a randomly selected angle $[0 .. 2\pi)$

X is the waiting time until the next packet arrives

The new → bit

can't put non-zero probability *at* points; *probability density function* assigns how probability mass is distributed *near* points; probability per unit length

$f(x): R \to R$, the *probability density function* (or simply "density")



Require:

$f(x) \geq 0$, and

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

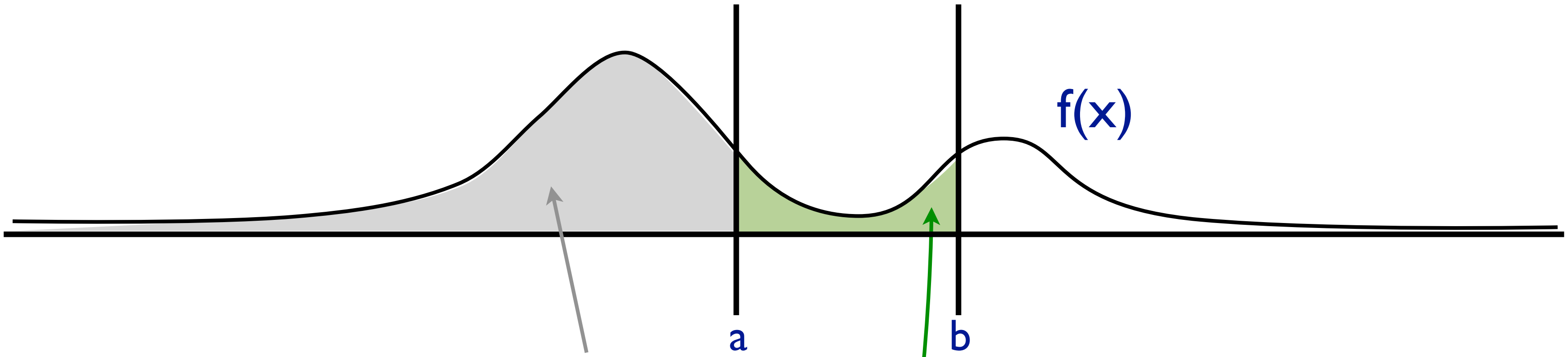
I.e., distribution is:

← nonnegative, and

← normalized,

just like discrete PMF. Distributes probability across the real line

F(x): the *cumulative distribution function* (aka the "distribution")



$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$ (Area left of a)

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

A key relationship:

$f(x) = \frac{d}{dx} F(x)$, since $F(a) = \int_{-\infty}^{a} f(x)\,dx$, Densities are *not* probabilities; e.g. may be > 1

$P(X = a) = \lim_{\varepsilon \to 0} P(a - \varepsilon/2 < X \le a + \varepsilon/2) = F(a) - F(a) = 0$

I.e.,

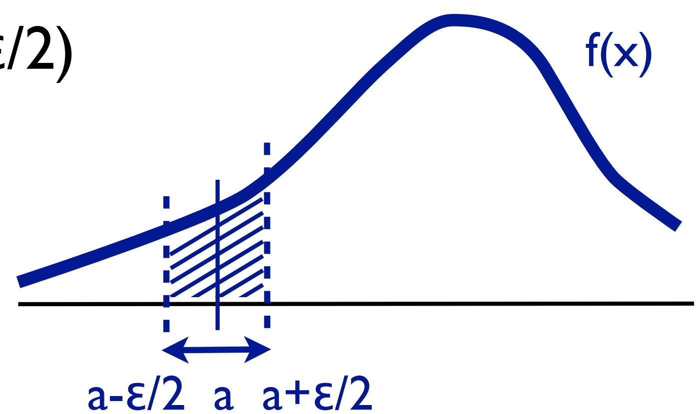
the probability that a continuous r.v. falls *at* a specified point is *zero*.

But

the probability that it falls *near* that point is *proportional to the density*:

$P(a - \varepsilon/2 < X \le a + \varepsilon/2) = F(a + \varepsilon/2) - F(a - \varepsilon/2)$

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



I.e.,

- $f(a) \approx$ *probability per unit length* near a.
- in a large random sample, expect more samples where density is higher (hence the name "density").
- $f(a)$ vs $f(b)$ give *relative probabilities* near a vs b.

Much of what we did with discrete r.v.s carries over almost unchanged, with

(a) p.m.f. replaced by *density*, and

(b) $\Sigma_x\ldots$ replaced by $\int \ldots dx$

E.g.

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

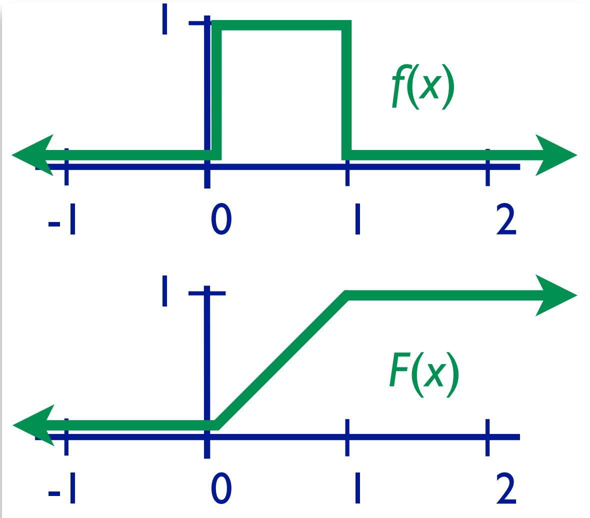
For continuous r.v. X, $E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$

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 $\approx \Sigma\, xf(x)dx$ (summed over grid points spaced dx apart on the real line) and the limit of that as $dx \to 0$ is $\int xf(x)dx$

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



$$
\begin{aligned}
F(a) &= \int_{-\infty}^{a} f(x)dx \\
&= \begin{cases} 0 & \text{if } a \le 0 \\ a & \text{if } 0 < a \le 1 \text{ (since } a = \int_0^a 1dx) \\ 1 & \text{if } 1 < a \end{cases}
\end{aligned}
$$

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

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

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

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

Alternatively, let $Y = g(X)$, find the density of Y, say f_Y, (see B&T 4.1; somewhat like r.v. slides 33-35) and directly compute $E[Y] = \int y f_Y(y) dy$.

Definition is same as in the discrete case

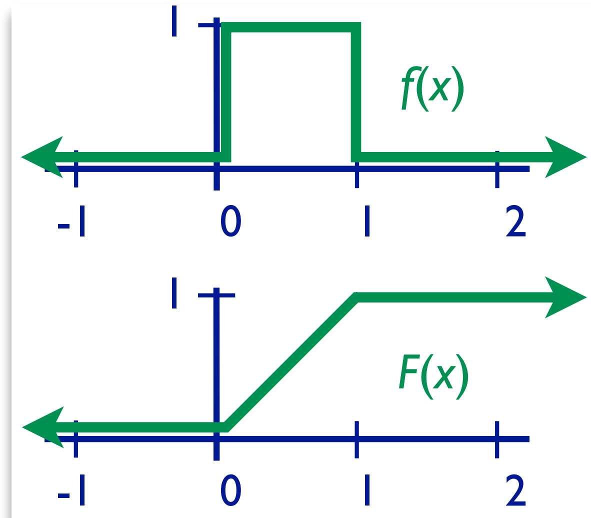
$\mathrm{Var}[X] = E[(X-\mu)^2]$ where $\mu = E[X]$

Identity still holds:

$\mathrm{Var}[X] = E[X^2] - (E[X])^2$ proof "same"

example

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



$$
\begin{aligned}
F(a) &= \int_{-\infty}^{a} f(x)dx \\
&= \begin{cases} 0 & \text{if } a \le 0 \\ a & \text{if } 0 < a \le 1 \ (\text{since } a = \int_0^a 1dx) \\ 1 & \text{if } 1 < a \end{cases}
\end{aligned}
$$

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

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

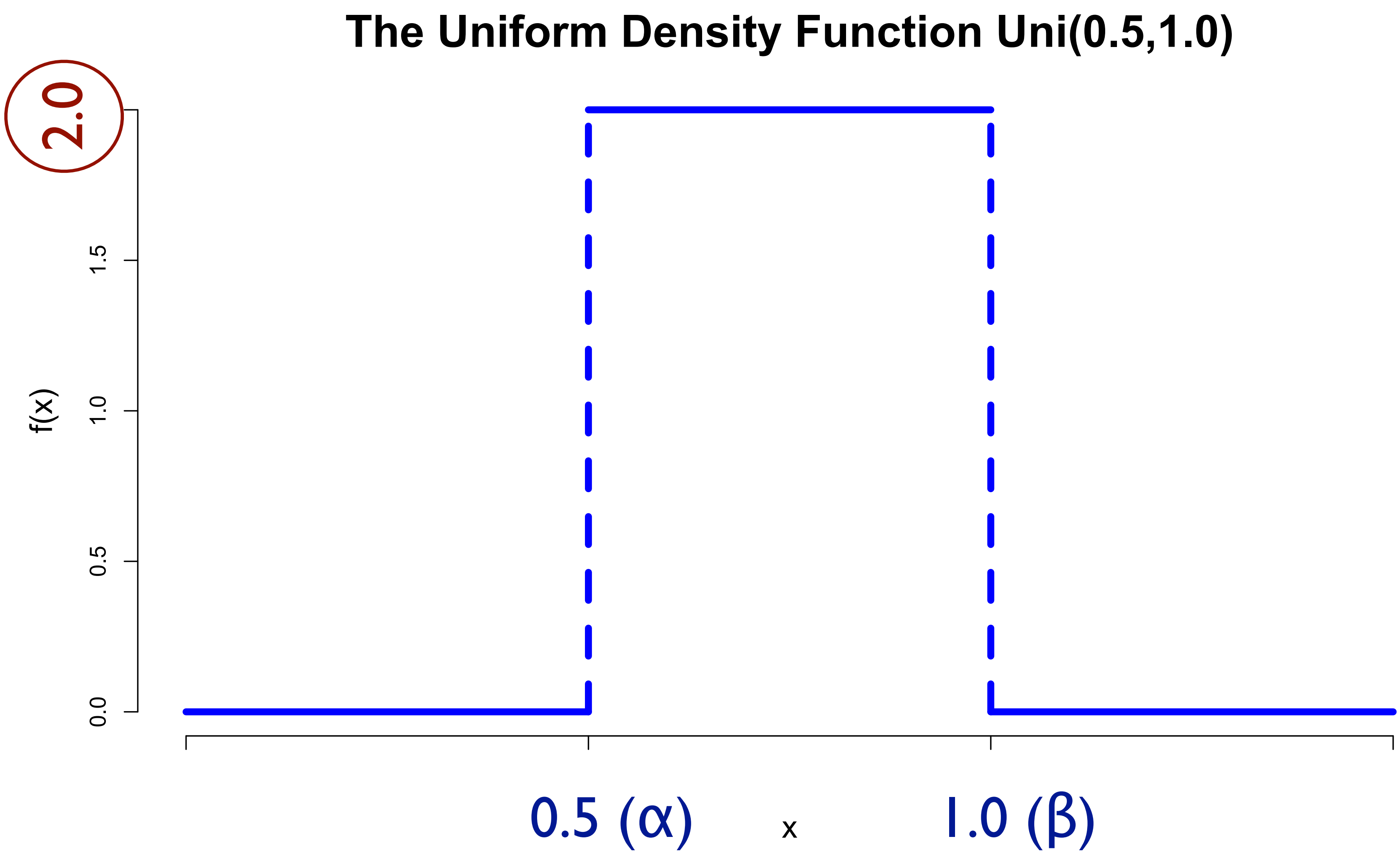
Continuous random variable X has density f(x), and

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

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

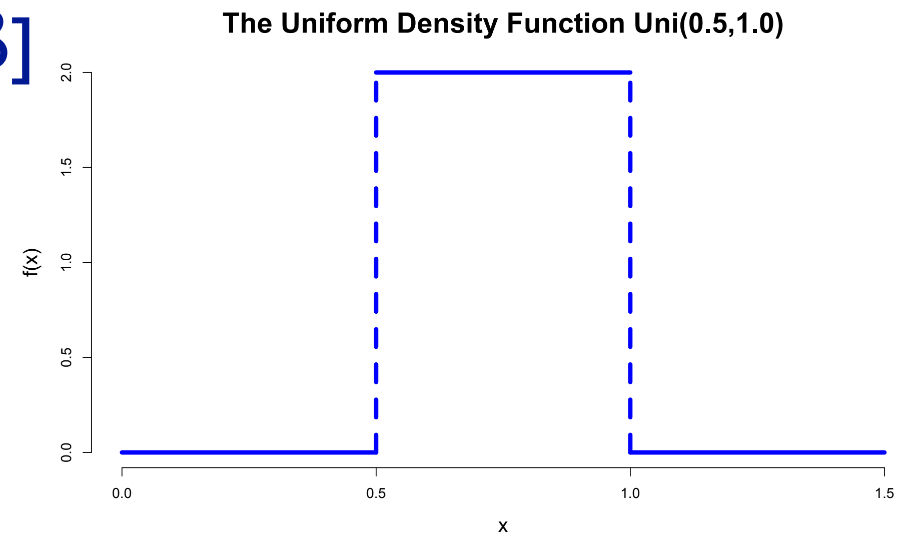
$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx$$

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



X ~ Uni(α,β) is uniform in [α,β]

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



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

if α≤a≤b≤β:

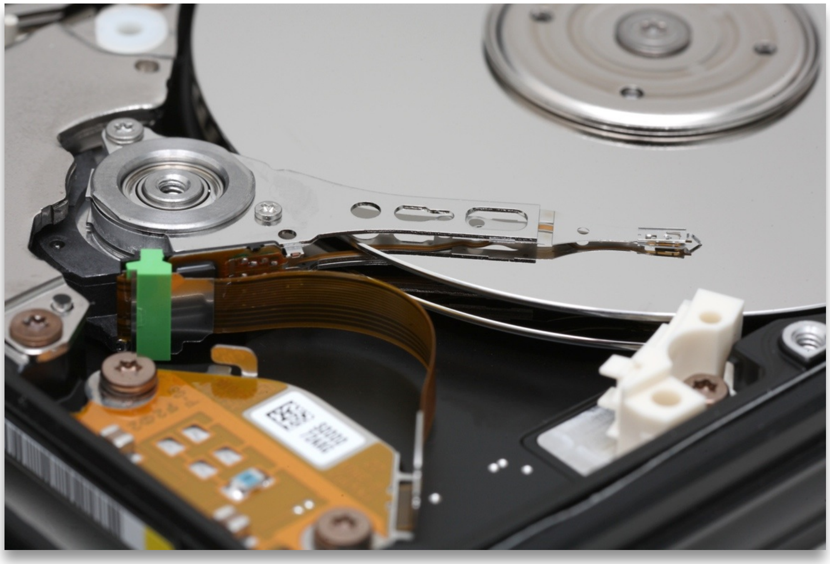
Yes, you should review your basic calculus; e.g., these 2 integrals would be good practice.

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

X ~ Uni(α,β) is uniform in [α,β]

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

You want to read a disk sector from a 7200rpm disk drive. Let T be the time you wait, in milliseconds, after the disk head is positioned over the correct track, until the desired sector rotates under the head.



T ~ Uni(0, 8.33)

Average Wait? 4.17ms

Radioactive decay: How long until the next alpha particle?

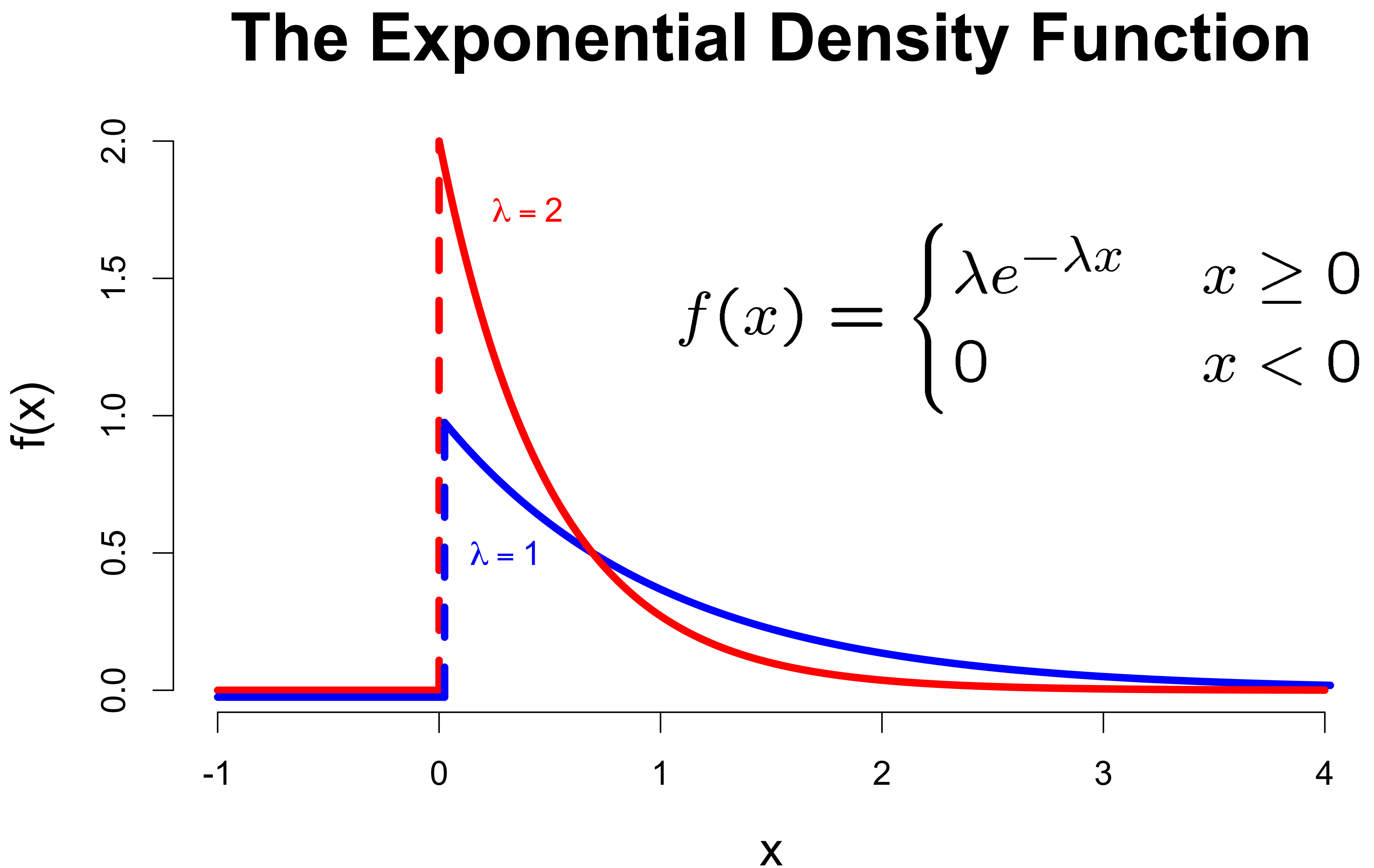
Customers: how long until the next customer/packet arrives at the checkout stand/server?

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?

Assuming events are independent, happening at some fixed *average* rate of λ per unit time – the waiting time until the next event is exponentially distributed (next slide)

X ~ Exp(λ)


The Exponential Density Function
λ = 2
λ = 1
$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
f(x)
x
2.0
1.5
1.0
0.5
0.0
-1
0
1
2
3
4

X ~ Exp(λ)

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

$$E[X] = \frac{1}{\lambda} \qquad \mathrm{Var}[X] = \frac{1}{\lambda^2}$$

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

Memorylessness:

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

Assuming exp distr, if you've waited *s* minutes, prob of waiting *t* more is exactly same as *s = 0*

Gambler's fallacy: "I'm due for a win"

Relation to the Poisson: same process, different measures:

Poisson: *how many* events in a *fixed time*;
Exponential: *how long* until the *next event*

λ is avg # per unit time;
$1/\lambda$ is mean wait

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, vs
How long to a Head, 3 flips per sec, prob p/3, vs
⋮

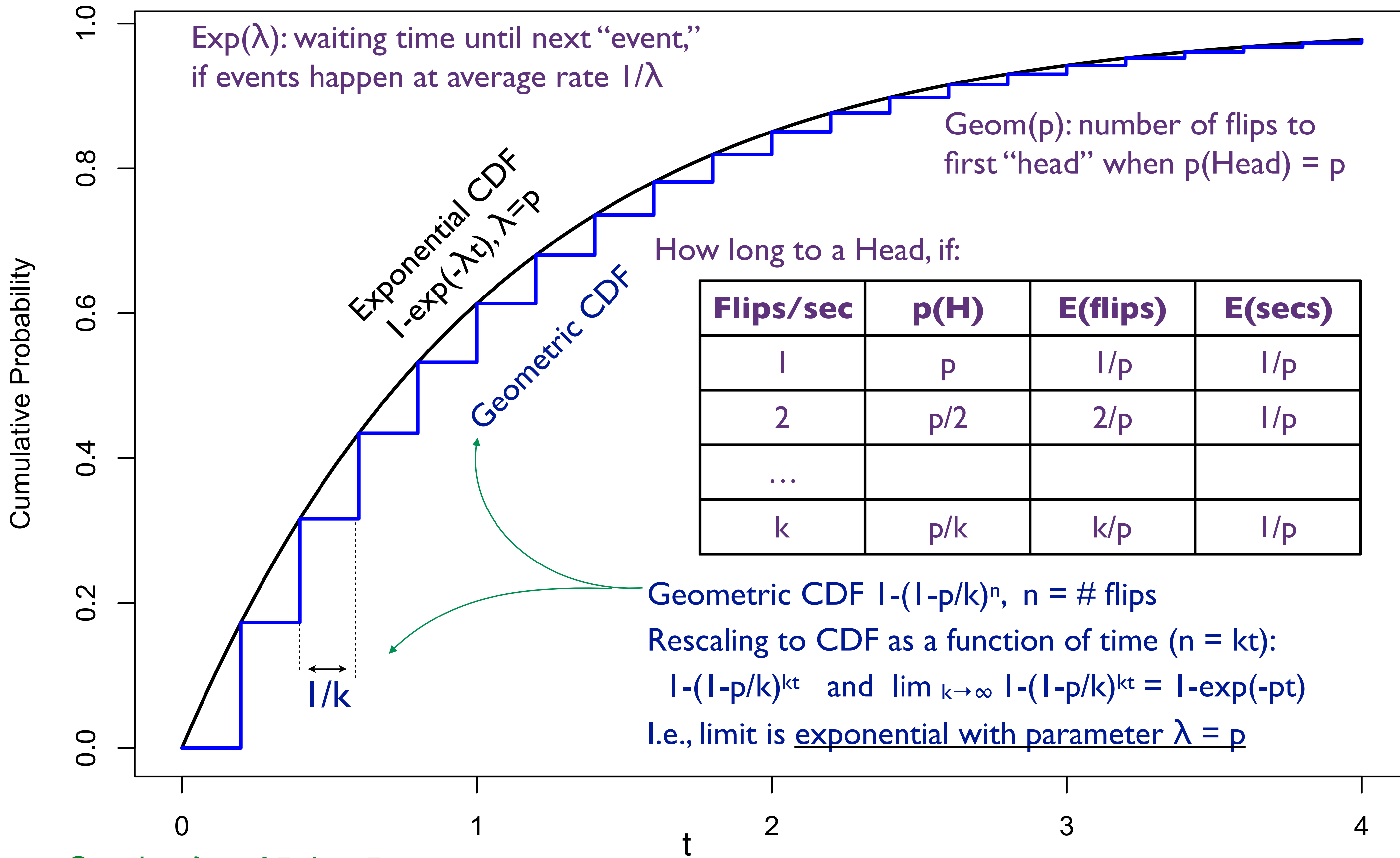
All have the same mean: 1/p secs

Limit is exponential with parameter 1/p

see also B&T fig 3.8, p152

geometric is discrete analog of exponential

cf also B&T fig 3.8, pg 152



Exp(λ): waiting time until next "event," if events happen at average rate 1/λ

Geom(p): number of flips to first "head" when p(Head) = p

How long to a Head, if:

Flips/sec	p(H)	E(flips)	E(secs)
1	p	1/p	1/p
2	p/2	2/p	1/p
...			
k	p/k	k/p	1/p

Geometric CDF $1-(1-p/k)^n$, n = # flips

Rescaling to CDF as a function of time (n = kt):

$1-(1-p/k)^{kt}$ and $\lim_{k\to\infty} 1-(1-p/k)^{kt} = 1-\exp(-pt)$

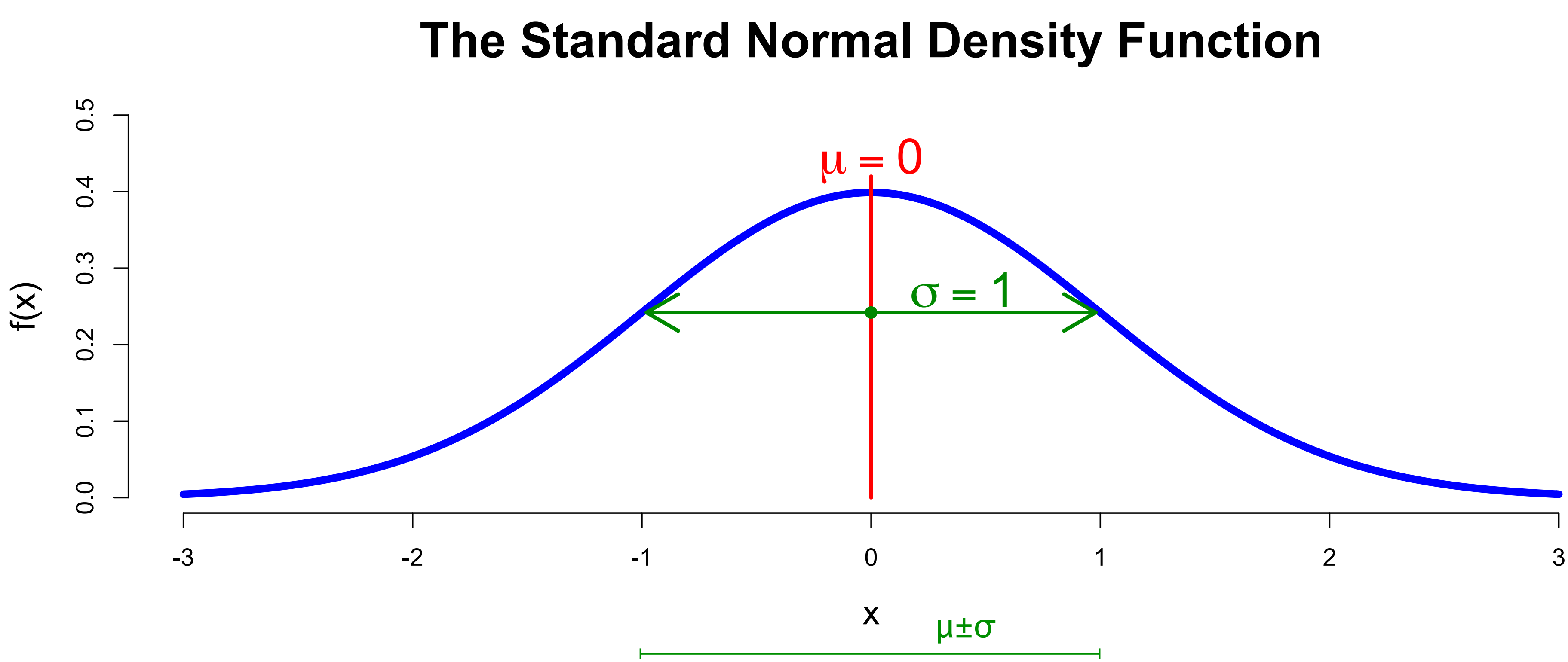
I.e., limit is <u>exponential with parameter λ = p</u>

Graphs: λ = .95, k = 5

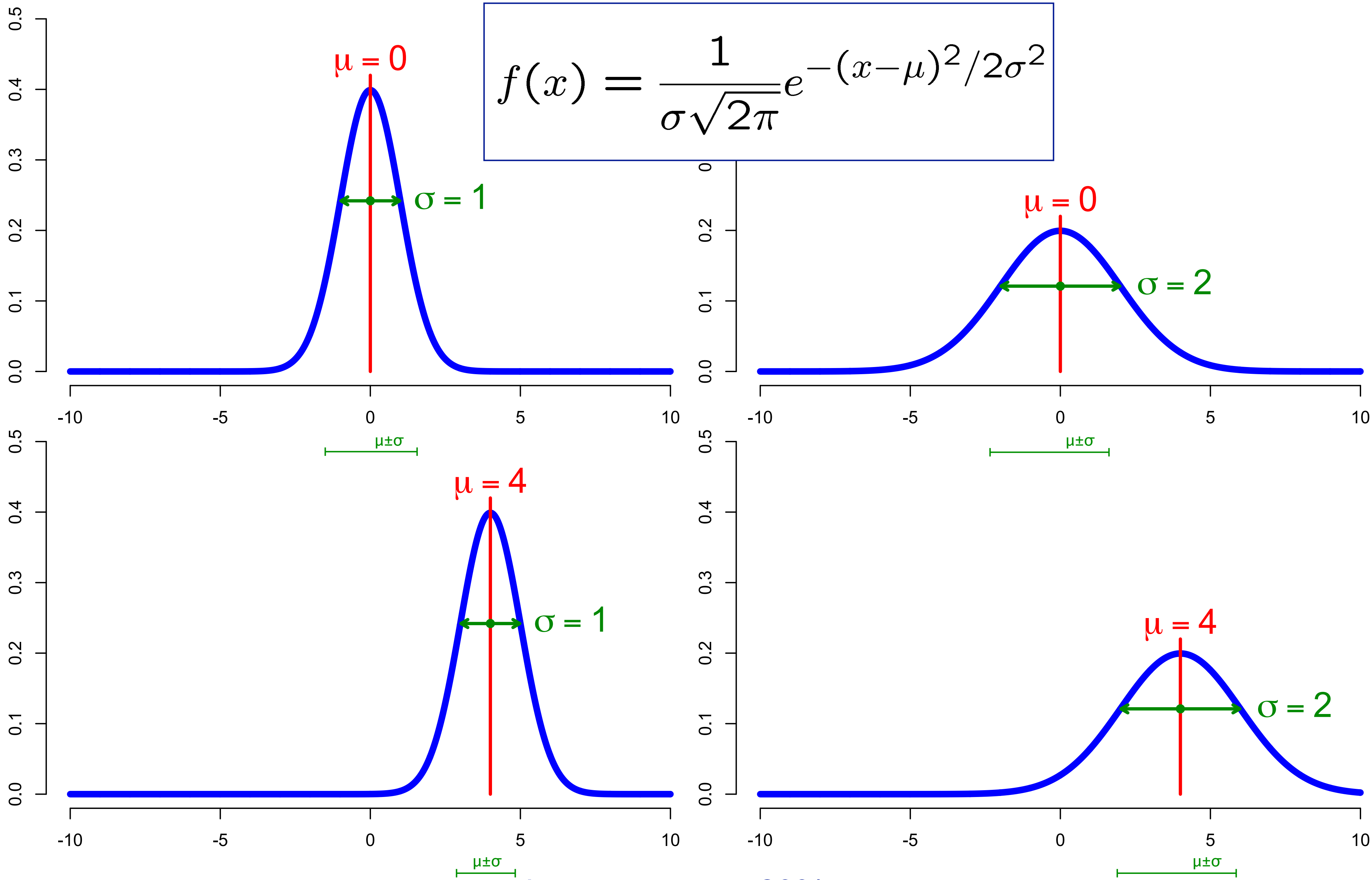
X is a normal (aka Gaussian) random variable $X \sim N(\mu, \sigma^2)$

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

$$E[X] = \mu \qquad \mathrm{Var}[X] = \sigma^2$$



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



density at μ is ≈ .399/σ

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

$Y = aX + b$

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

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

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

E[ ], Var[ ] as expected;
"normality" is the surprise

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

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

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

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

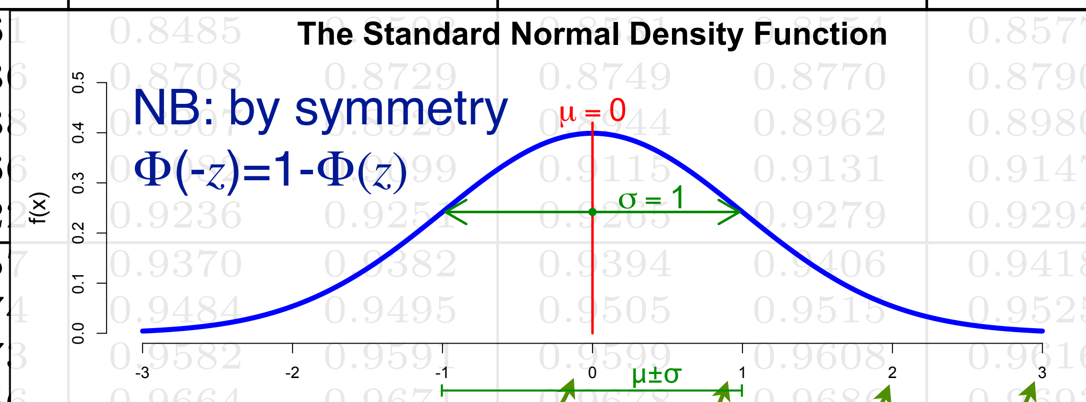
$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

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	0.7123	[illegible]	0.7190
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	[illegible]	0.7517
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	[illegible]	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	[illegible]	[illegible]	[illegible]	0.8577	0.8599
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	[illegible]	[illegible]	[illegible]	0.8962	0.8980	0.8997
1.3	0.9032	0.9049	0.9066	[illegible]	[illegible]	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	0.9686	0.9693	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	[illegible]	[illegible]	[illegible]
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	[illegible]	[illegible]	[illegible]
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

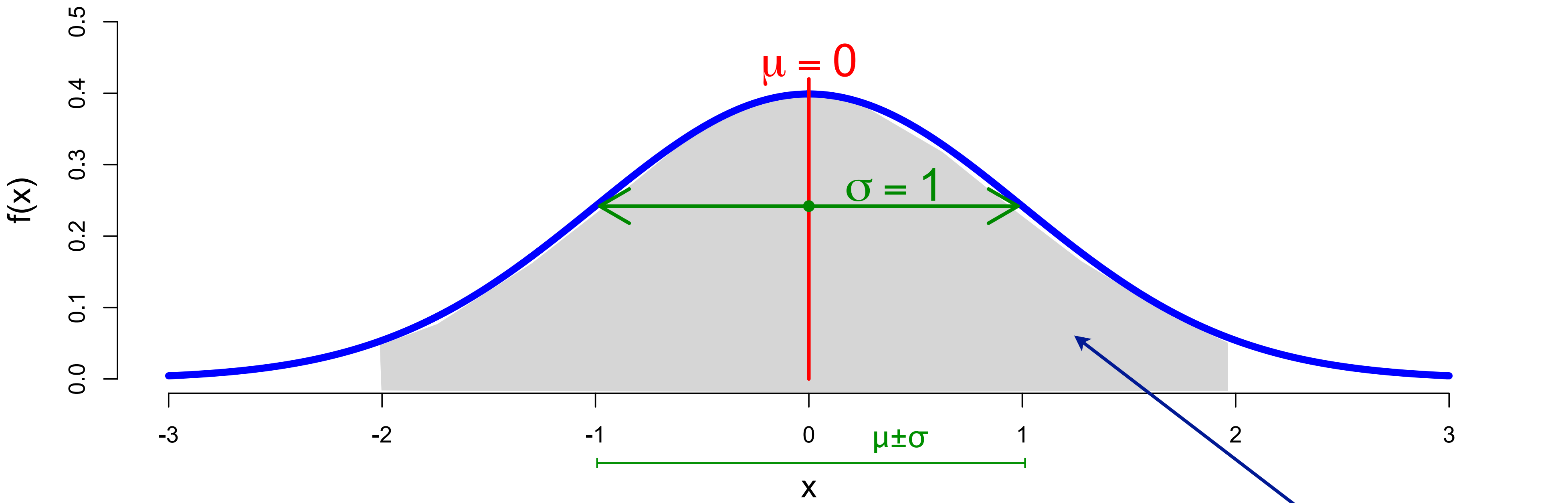
$\Phi(.46)$



NB: by symmetry $\Phi(-z)=1-\Phi(z)$

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

The Standard Normal Density Function



If $Z \sim N(\mu,\sigma^2)$ what is $P(\ \mu-\sigma < Z < \mu+\sigma\ )$?

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

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

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

Why?

$$\mu - k\sigma < \overset{N(\mu,\sigma^2)}{\boxed{Z}} < \mu + k\sigma \quad \Leftrightarrow \quad -k < \overset{N(0,1)}{\boxed{(Z-\mu)/\sigma}} < +k$$

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

X_i has $\mu = E[X_i] < \infty$ and $\sigma^2 = Var[X_i] < \infty$

As $n \to \infty$,

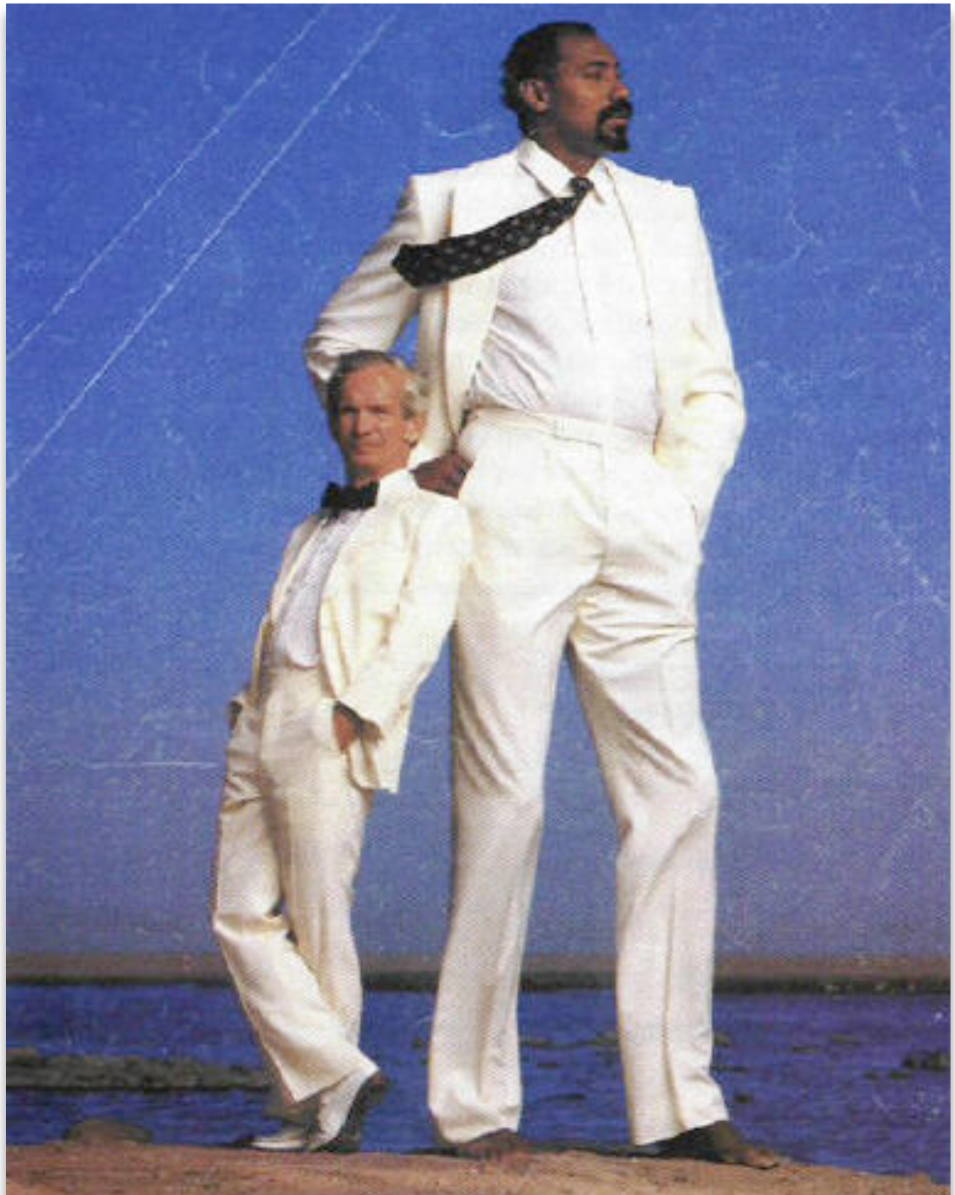
$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

Restated: As $n \to \infty$,

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

More of the theory behind this later, but first, some examples:

How tall are you? Why?



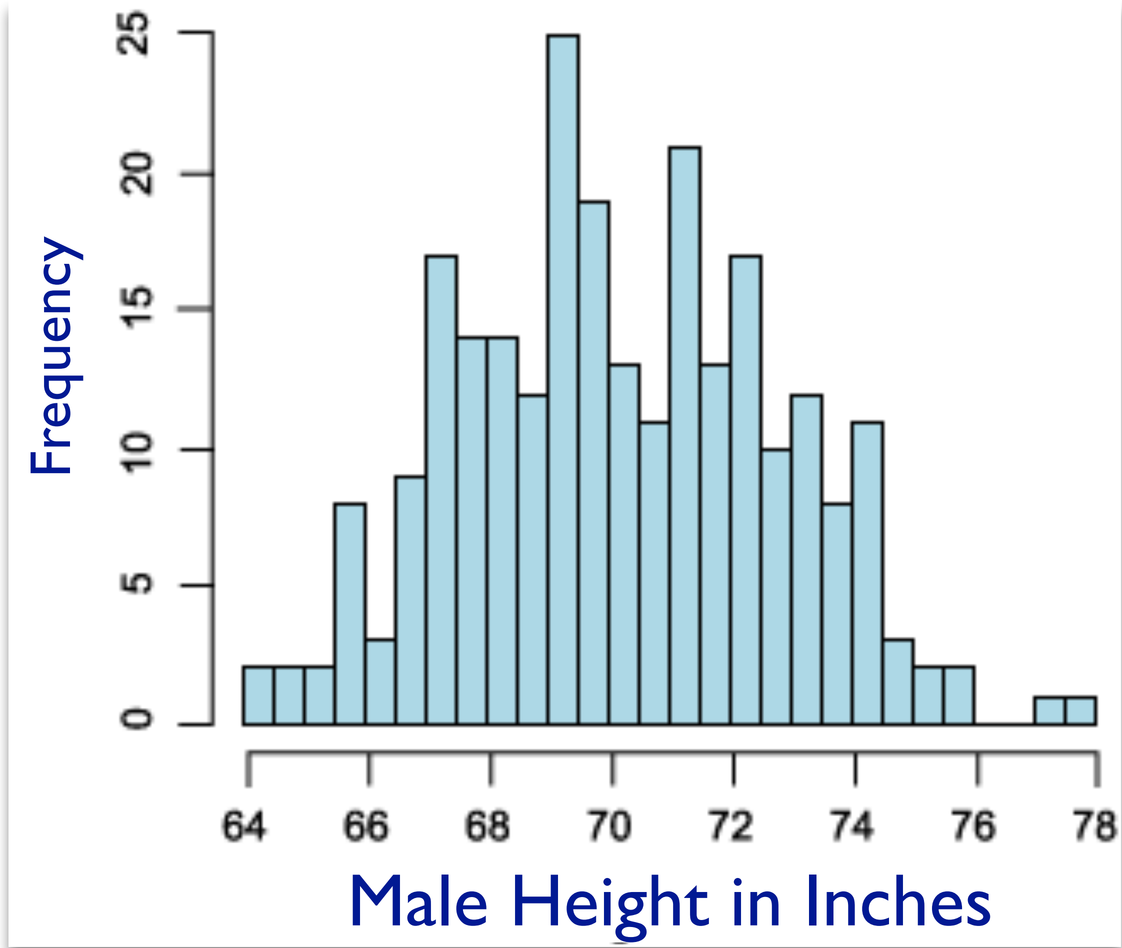
Credit: Annie Leibovitz, © 1987 ?

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, DNA, etc. were...)



Meta-analysis of Dense Genecentric Association Studies Reveals Common and Uncommon Variants Associated with Height, Lanktree, et al. 194
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

(and hundreds more probably exist)

Locus Rank	Chr.	Candidate Gene[a]	SNP[a]	Effect Allele	MAF	European Phase I (up to 5 Effect
1	7q22	*CDK6*	rs4272	A	0.21	−0.46
2	6p21	*HMGA1*	rs1150781	C	0.09	0.73
3	12q15	*HMGA2*				
4	20q11	*MMP24*				
5	17q23	*MAP3K3*				
6	17q24	*GH1-GH2*				
7	1p36	*MFAP2*				
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]	Ef Al
28	2p23	*GCKR*	rs780094	T
29	1q41	*TGFB2*	rs900	A
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*		

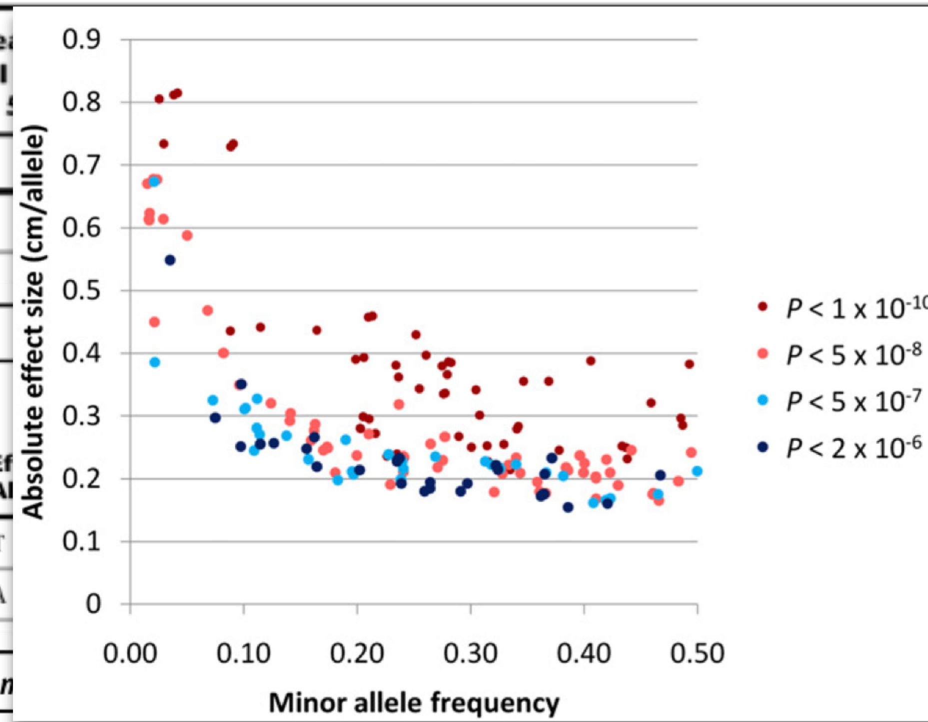
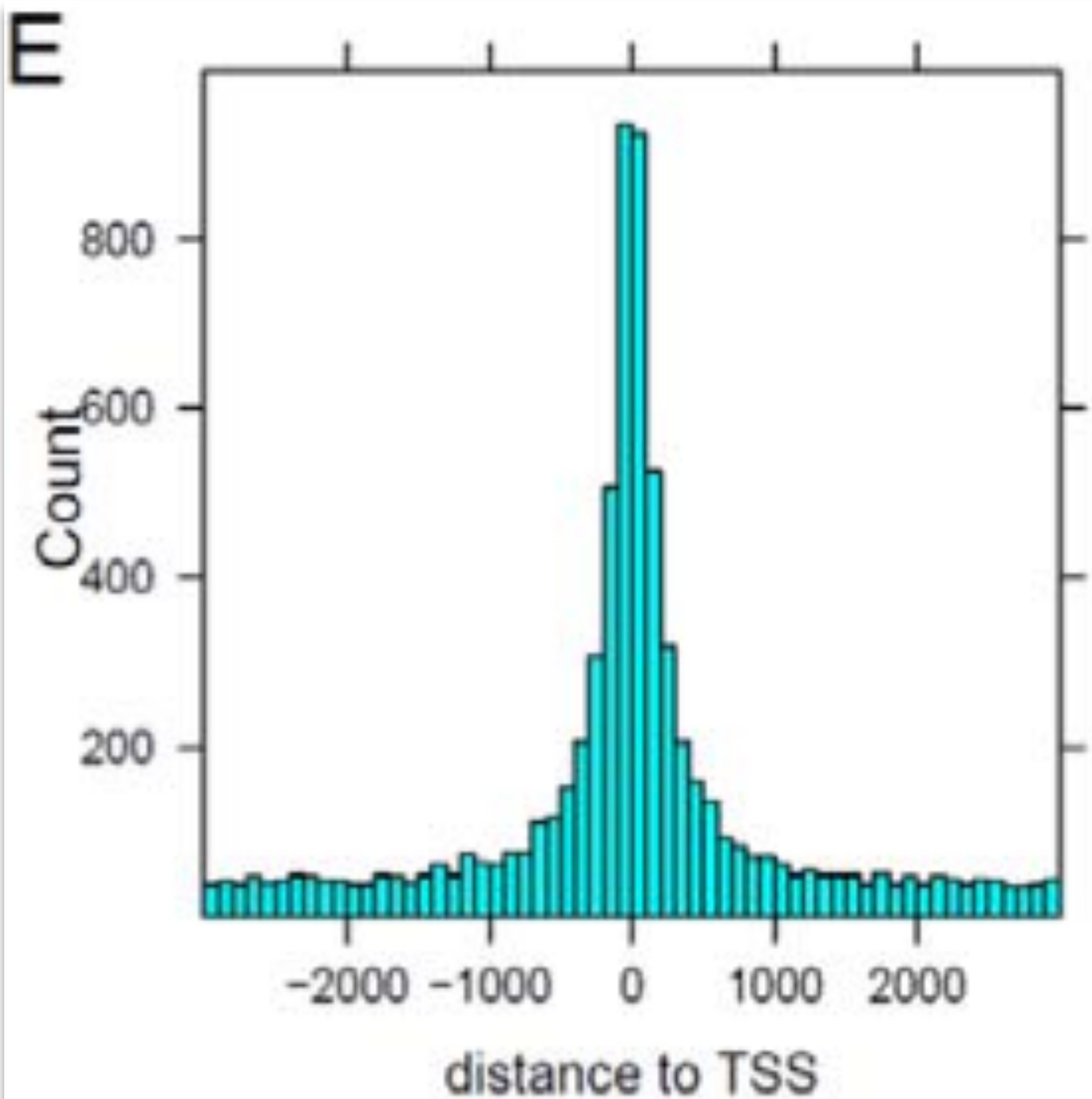


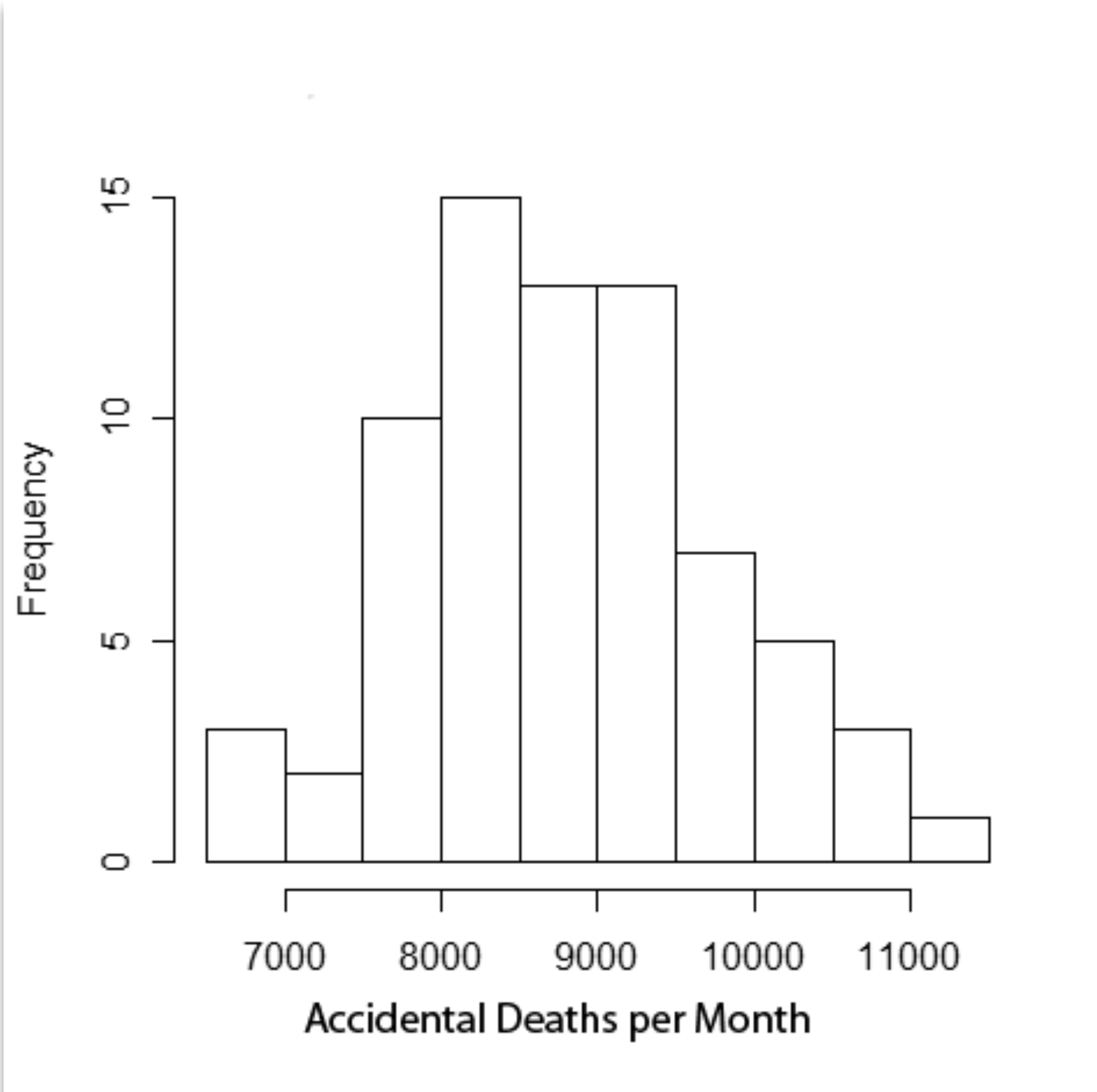
Table 1. Con

Locus Rank	Chr.	Candidate Gene[a]	SNP[a]	Effect Allele	MAF	European Ancestry Phase I (up to 53,394) Effect	p	r²
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

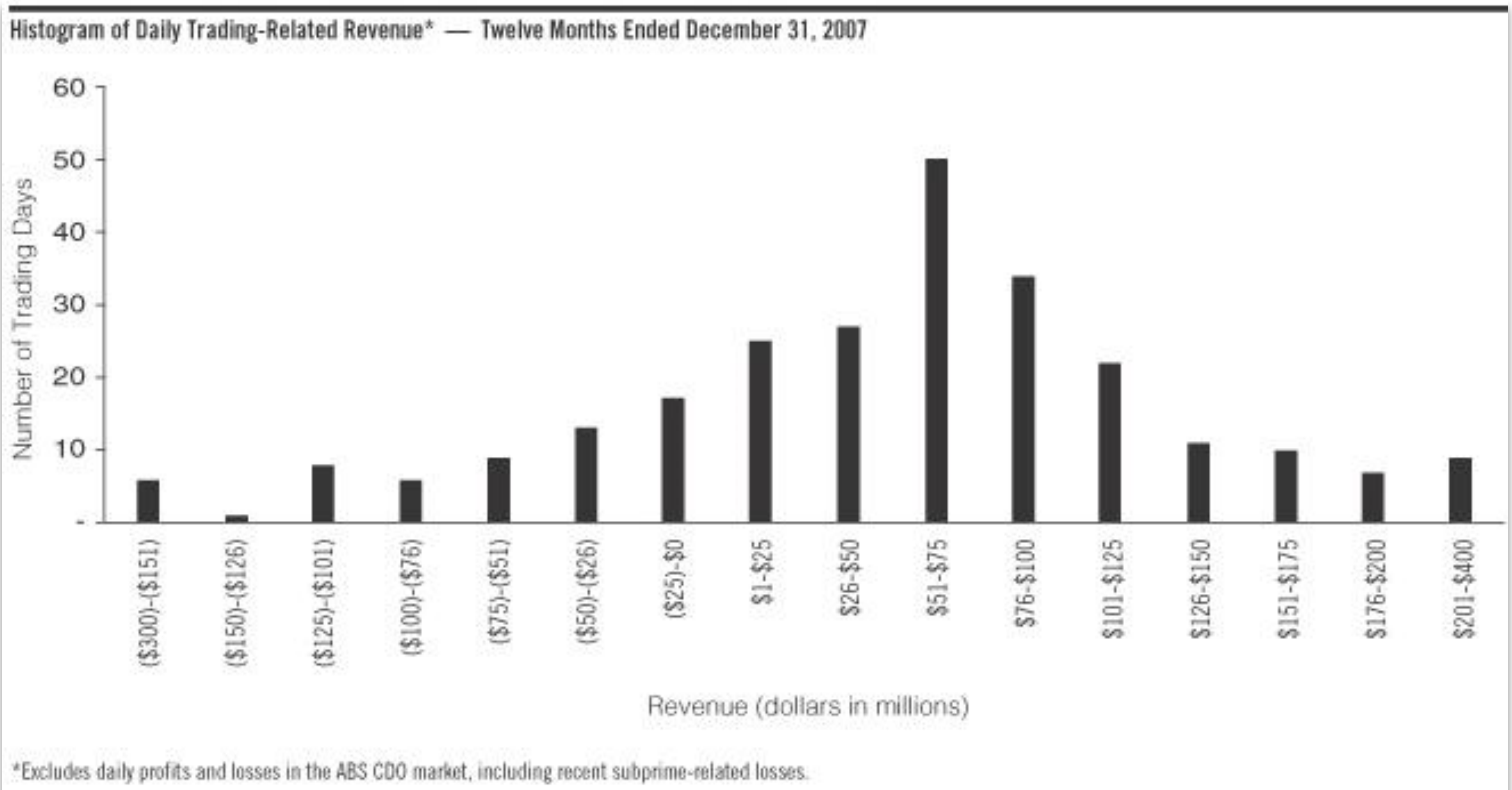
A Single *IGF1* Allele Is a Major Determinant of Small Size in Dogs Nathan B. Sutter, *et al. Science* **316**, 112 (2007);


E
Count
800
600
400
200
-2000
-1000
0
1000
2000
distance to TSS

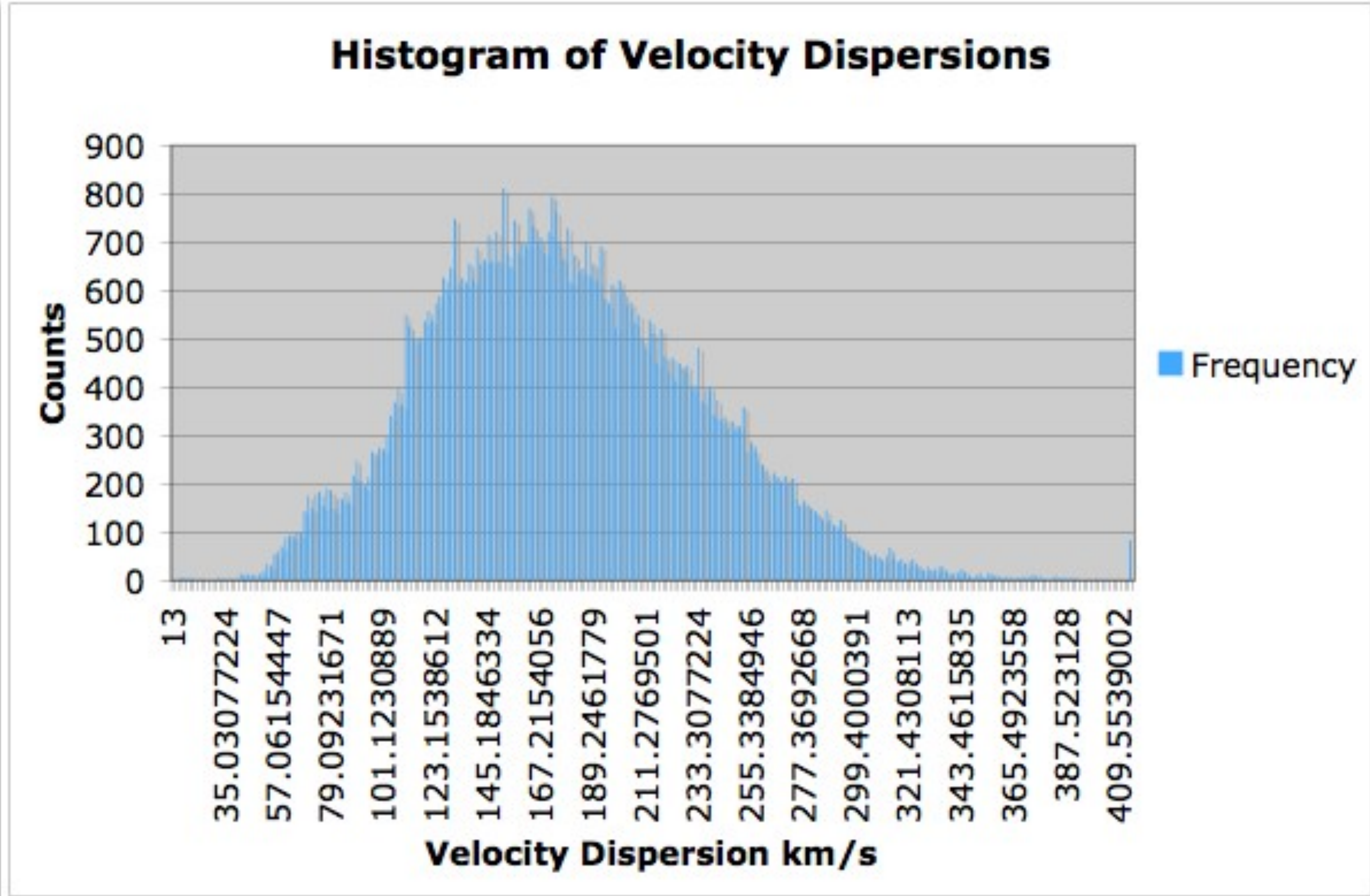
in the real world…



in the real world…



in the real world…



pdf vs cdf

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

sums become integrals, e.g.

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

most familiar properties still hold, e.g.

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

$Var[X] = E[X^2] - (E[X])^2$

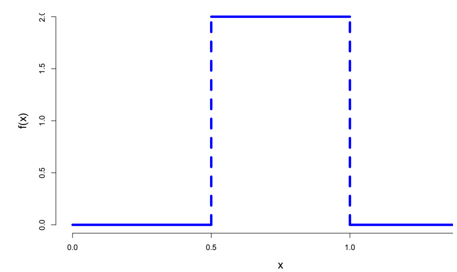
Three important examples

X ~ Uni(α,β) uniform in [α,β]

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

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

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

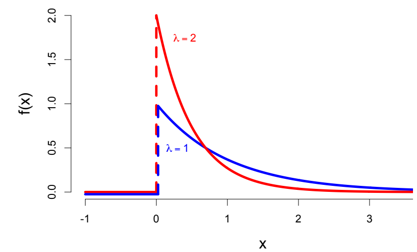


X ~ Exp(λ) exponential

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

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

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

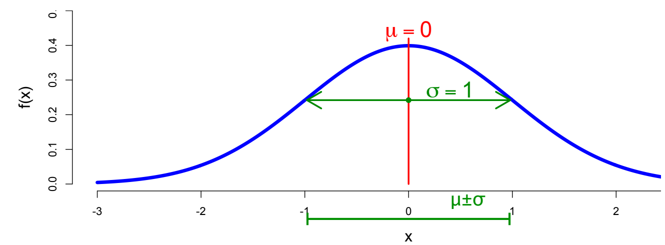


X ~ N(μ, σ²) normal (aka Gaussian, aka the big Kahuna)

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

$$E[X] = \mu$$

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



Joint, marginal and conditional distributions

Distribution of $Z = g(X,Y)$ based on distributions of X,Y

Independence

- Joint = product of marginals, and/or
- Conditional = unconditional

Chain rule

Total probability rule

Conditional expectation

Law of total expectation

…

See Text, Chapter 3!