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

\[ X \in \{1, 2, \ldots, 6\} \] 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): $\mathbb{R} \to \mathbb{R}$, the *probability density function* (or simply “density”)

Require:

\[ f(x) \geq 0, \text{ and} \]

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

I.e., distribution is:

\[ \leftarrow \text{nonnegative, and} \]

\[ \leftarrow \text{normalized,} \]

just like discrete PMF
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:

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

\[ P(X = a) = \lim_{\varepsilon \to 0} P(a-\varepsilon < X \leq a) = F(a) - F(a) = 0 \]

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**:

\[ P(a - \varepsilon/2 < X \leq a + \varepsilon/2) = \]

\[ F(a + \varepsilon/2) - F(a - \varepsilon/2) \]

\[ \approx \varepsilon \cdot f(a) \]

I.e., 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 $\sum_x \ldots$ replaced by $\int \ldots \, dx$

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 $\approx \sum xf(x)dx$ (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 \( f(x) = \begin{cases} 
1 & \text{for } 0 < x < 1 \\
0 & \text{elsewhere} 
\end{cases} \)

\[
F(a) = \int_{-\infty}^{a} f(x) \, dx
\]

\[
= \begin{cases} 
0 & \text{if } a \leq 0 \\
a & \text{if } 0 < a \leq 1 \text{ (since } a = \int_{0}^{a} 1 \, dx) \\
1 & \text{if } 1 < a 
\end{cases}
\]

\[
E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{1} x \, dx = \frac{x^2}{2} \bigg|_{0}^{1} = \frac{1}{2}
\]

\[
E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) \, dx = \int_{0}^{1} x^2 \, dx = \frac{x^3}{3} \bigg|_{0}^{1} = \frac{1}{3}
\]

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

\[ E[aX+b] = aE[X]+b \]
\[ E[X+Y] = E[X]+E[Y] \]

Functions of a random variable

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

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 \( E[Y] = \int yf_Y(y)dy \).
Definition is same as in the discrete case

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

Identity still holds:

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

proof “same”
Let \( f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases} \)

\[
F(a) = \int_{-\infty}^{a} f(x) \, dx
\]

\[
= \begin{cases} 0 & \text{if } a \leq 0 \\ a & \text{if } 0 < a \leq 1 \quad (\text{since } a = \int_{0}^{a} 1 \, dx) \\ 1 & \text{if } 1 < a \end{cases}
\]

\[
E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{1} x \, dx = \frac{x^2}{2} \bigg|_{0}^{1} = \frac{1}{2}
\]

\[
E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) \, dx = \int_{0}^{1} x^2 \, dx = \frac{x^3}{3} \bigg|_{0}^{1} = \frac{1}{3}
\]

\[
\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \quad (\sigma \approx 0.29)
\]
Continuous random variable $X$ has density $f(x)$, and

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

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

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x) \, dx$$
$X \sim \text{Uni}(\alpha, \beta)$ is uniform in $[\alpha, \beta]$ 

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

The Uniform Density Function Uni(0.5, 1.0)
$X \sim \text{Uni}(\alpha, \beta)$ 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 $\alpha \leq a \leq b \leq \beta$:

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

Yes, you should review your basic calculus; e.g., these 2 integrals would be good practice.
\( X \sim \text{Uni}(\alpha, \beta) \) is uniform in \([\alpha, \beta]\)

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

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.

\( T \sim \text{Uni}(0, 8.33) \)

Average Wait? 4.17ms
waiting for “events”

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-carrages, etc., can they stick to it?

Assuming events are independent, happening at some fixed average rate of $\lambda$ per unit time – the waiting time until the next event is exponentially distributed (next slide)
$X \sim \text{Exp}(\lambda)$

The Exponential Density Function

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

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

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

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

Memorylessness:

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

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

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

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
⋮
Limit is exponential with parameter 1/p

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

See also B&T fig 3.8, p152
geometric is discrete analog of exponential

Exp(\(\lambda\)): waiting time until next “event,” if events happen at average rate \(1/\lambda\)

Geom(p): number of flips to first “head” when \(p(\text{Head}) = p\)

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

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 exponential with parameter \(\lambda = p\)

Graphs: \(\lambda = .95\), \(k = 5\)

cf also B&T fig 3.8, pg 152
The Standard Normal Density Function

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

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

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

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

density at $\mu$ is $\approx .399/\sigma$
**normal random variables**

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

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

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

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

no closed form 😞
Table of the Standard Normal Cumulative Distribution Function $\Phi(z)$

<table>
<thead>
<tr>
<th>$z$</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7485</td>
<td>0.7517</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8463</td>
<td>0.8488</td>
<td>0.8513</td>
<td>0.8538</td>
<td>0.8562</td>
<td>0.8587</td>
<td>0.8611</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8665</td>
<td>0.8688</td>
<td>0.8708</td>
<td>0.8732</td>
<td>0.8757</td>
<td>0.8780</td>
<td>0.8803</td>
<td>0.8826</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8908</td>
<td>0.8929</td>
<td>0.8950</td>
<td>0.8971</td>
<td>0.8991</td>
<td>0.9011</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9065</td>
<td>0.9080</td>
<td>0.9095</td>
<td>0.9110</td>
<td>0.9125</td>
<td>0.9140</td>
<td>0.9154</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9249</td>
<td>0.9262</td>
<td>0.9275</td>
<td>0.9288</td>
<td>0.9300</td>
</tr>
<tr>
<td>1.5</td>
<td>0.9332</td>
<td>0.9345</td>
<td>0.9357</td>
<td>0.9367</td>
<td>0.9377</td>
<td>0.9385</td>
<td>0.9393</td>
<td>0.9401</td>
<td>0.9409</td>
</tr>
<tr>
<td>1.6</td>
<td>0.9452</td>
<td>0.9463</td>
<td>0.9473</td>
<td>0.9482</td>
<td>0.9491</td>
<td>0.9499</td>
<td>0.9507</td>
<td>0.9515</td>
<td>0.9523</td>
</tr>
<tr>
<td>1.7</td>
<td>0.9554</td>
<td>0.9564</td>
<td>0.9573</td>
<td>0.9581</td>
<td>0.9589</td>
<td>0.9597</td>
<td>0.9605</td>
<td>0.9612</td>
<td>0.9619</td>
</tr>
<tr>
<td>1.8</td>
<td>0.9641</td>
<td>0.9649</td>
<td>0.9656</td>
<td>0.9663</td>
<td>0.9670</td>
<td>0.9676</td>
<td>0.9683</td>
<td>0.9689</td>
<td>0.9695</td>
</tr>
<tr>
<td>1.9</td>
<td>0.9713</td>
<td>0.9719</td>
<td>0.9726</td>
<td>0.9732</td>
<td>0.9738</td>
<td>0.9744</td>
<td>0.9750</td>
<td>0.9756</td>
<td>0.9761</td>
</tr>
<tr>
<td>2.0</td>
<td>0.9772</td>
<td>0.9778</td>
<td>0.9783</td>
<td>0.9788</td>
<td>0.9793</td>
<td>0.9798</td>
<td>0.9803</td>
<td>0.9808</td>
<td>0.9812</td>
</tr>
<tr>
<td>2.1</td>
<td>0.9821</td>
<td>0.9826</td>
<td>0.9830</td>
<td>0.9834</td>
<td>0.9838</td>
<td>0.9842</td>
<td>0.9846</td>
<td>0.9850</td>
<td>0.9854</td>
</tr>
<tr>
<td>2.2</td>
<td>0.9861</td>
<td>0.9864</td>
<td>0.9868</td>
<td>0.9871</td>
<td>0.9875</td>
<td>0.9878</td>
<td>0.9881</td>
<td>0.9884</td>
<td>0.9887</td>
</tr>
<tr>
<td>2.3</td>
<td>0.9893</td>
<td>0.9896</td>
<td>0.9898</td>
<td>0.9901</td>
<td>0.9904</td>
<td>0.9906</td>
<td>0.9909</td>
<td>0.9911</td>
<td>0.9913</td>
</tr>
<tr>
<td>3.0</td>
<td>0.9987</td>
<td>0.9987</td>
<td>0.9987</td>
<td>0.9987</td>
<td>0.9988</td>
<td>0.9988</td>
<td>0.9989</td>
<td>0.9989</td>
<td>0.9990</td>
</tr>
<tr>
<td>3.1</td>
<td>0.9990</td>
<td>0.9991</td>
<td>0.9991</td>
<td>0.9992</td>
<td>0.9992</td>
<td>0.9992</td>
<td>0.9993</td>
<td>0.9993</td>
<td>0.9993</td>
</tr>
<tr>
<td>3.2</td>
<td>0.9993</td>
<td>0.9993</td>
<td>0.9994</td>
<td>0.9994</td>
<td>0.9994</td>
<td>0.9994</td>
<td>0.9994</td>
<td>0.9995</td>
<td>0.9995</td>
</tr>
<tr>
<td>3.3</td>
<td>0.9995</td>
<td>0.9995</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9996</td>
</tr>
<tr>
<td>3.4</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
</tr>
</tbody>
</table>

**The Standard Normal Density Function**

$\Phi(-z)=1-\Phi(z)$

**NB:** by symmetry

$\mu = 0$

$\sigma = 1$

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

$\mu - k\sigma < Z < \mu + k\sigma \iff -k < (Z-\mu)/\sigma < +k$
the central limit theorem (CLT)

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

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

As \( n \to \infty \),

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

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?

Willie Shoemaker & Wilt Chamberlain

Credit: Annie Leibovitz, © 1987
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...)

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

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

<table>
<thead>
<tr>
<th>Locus Rank</th>
<th>Chr.</th>
<th>Candidate Gene*</th>
<th>SNP*</th>
<th>Effect Allele</th>
<th>MAF</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7q22</td>
<td>CDK6</td>
<td>rs4272</td>
<td>A</td>
<td>0.21</td>
<td>−0.46</td>
</tr>
<tr>
<td>2</td>
<td>6p21</td>
<td>HMGAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12q15</td>
<td>HMGAI2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20q11</td>
<td>MMP24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>17q23</td>
<td>MAP3K3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>17q24</td>
<td>GH1-GH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1p36</td>
<td>MFAP2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>15q26</td>
<td>IGFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>7p22</td>
<td>GNA12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>17q23</td>
<td>TBX2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>12q22</td>
<td>SOCS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>9q22</td>
<td>PTCH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>14q11</td>
<td>NFATC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>15q26</td>
<td>ACAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2q24</td>
<td>NPPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>6p21</td>
<td>PPARD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>20q11</td>
<td>MYH7B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>19q13</td>
<td>IL11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sixty-Four** (and hundreds more probably exist)

Table 1. Continued

<table>
<thead>
<tr>
<th>Locus Rank</th>
<th>Chr.</th>
<th>Candidate Gene*</th>
<th>SNP*</th>
<th>Effect Allele</th>
<th>MAF</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>2p23</td>
<td>GCKR</td>
<td>rs780094</td>
<td>T</td>
<td>0.15</td>
<td>0.72</td>
</tr>
<tr>
<td>29</td>
<td>1q41</td>
<td>TGFBI</td>
<td>rs900</td>
<td>T</td>
<td>0.15</td>
<td>0.72</td>
</tr>
<tr>
<td>30</td>
<td>20q11</td>
<td>CDK5RAP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>2p12</td>
<td>EIF2AK3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>19q13</td>
<td>INSR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>6q25</td>
<td>ESR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>2q37</td>
<td>DIS3L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>2q35</td>
<td>PLC4D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1p36</td>
<td>RPS6AK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>15q21</td>
<td>CYP19A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>5q31</td>
<td>SLC22A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>7p15</td>
<td>JAZF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>17p13</td>
<td>POLR2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1p22</td>
<td>PKN2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>7q22</td>
<td>CNOT4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Continued

<table>
<thead>
<tr>
<th>Locus Rank</th>
<th>Chr.</th>
<th>Candidate Gene*</th>
<th>SNP*</th>
<th>Effect Allele</th>
<th>MAF</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>5q21</td>
<td>PCRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1q43</td>
<td>TCF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>12p13</td>
<td>KLF24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>19p13</td>
<td>NME4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>2p12</td>
<td>ESR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>1p36</td>
<td>RPS6AK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>15q21</td>
<td>CYP19A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5q31</td>
<td>SLC22A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>7p15</td>
<td>JAZF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>17p13</td>
<td>POLR2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>1p22</td>
<td>PKN2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>1p22</td>
<td>COL24A1</td>
<td>rs2046159</td>
<td>A</td>
<td>0.16</td>
<td>0.23</td>
</tr>
<tr>
<td>55</td>
<td>1q23</td>
<td>DUSP23</td>
<td>rs1129923</td>
<td>A</td>
<td>0.10</td>
<td>−0.25</td>
</tr>
<tr>
<td>56</td>
<td>10q22</td>
<td>MAT1A</td>
<td>rs7087728</td>
<td>A</td>
<td>0.18</td>
<td>0.22</td>
</tr>
<tr>
<td>57</td>
<td>2p15</td>
<td>PPP3R1</td>
<td>rs1822469</td>
<td>T</td>
<td>0.41</td>
<td>−0.14</td>
</tr>
<tr>
<td>58</td>
<td>7q36</td>
<td>ATG9B</td>
<td>rs1800783</td>
<td>A</td>
<td>0.38</td>
<td>−0.16</td>
</tr>
<tr>
<td>59</td>
<td>14q11</td>
<td>BCL2L2</td>
<td>rs3210043</td>
<td>A</td>
<td>0.16</td>
<td>0.25</td>
</tr>
<tr>
<td>60</td>
<td>4p14</td>
<td>RFC1</td>
<td>rs11096991</td>
<td>T</td>
<td>0.35</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**European Ancestry Phase I (up to 53,394)**
in the real world...
in the real world…
in the real world…

Histogram of Daily Trading-Related Revenue* — Twelve Months Ended December 31, 2007

*Excludes daily profits and losses in the ABS CDO market, including recent subprime-related losses.
in the real world…
continuous r.v.’s: summary

pdf vs cdf

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

sums become integrals, e.g.

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

most familiar properties still hold, e.g.

$$E[aX+bY+c] = aE[X]+bE[Y]+c$$
$$Var[X] = E[X^2] - (E[X])^2$$
continuous r.v.’s: summary

Three important examples

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

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

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

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

\( X \sim \text{Exp}(\lambda) \) 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, aka the big Kahuna)

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

\( E[X] = \mu \)

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