Some Statistics Notes PhL 2.7-10.15

These notes are an addendum to already existing noted in **Ng Notes** and **Fourier Appendix G**.

A. Review of Ng notes and Throwing one Die 2

B. Theorems about Variance and Covariance 3

Theorems 1 through 6 4

Theorem 7 (Standard Deviation of the Mean Theorem) 7

C. The Binomial Distribution and the Bernoulli Trial 8

Theorem 8: (Bernoulli Trial SD of the Mean Theorem) 12

(a) I try but fail to show that Binomial Distribution approaches a Gaussian for large N 12

(b) Various Binomial Notations and Access from Excel or Maple 13

(c) The Central Limit Theorem and my Second Failed Attempt to show Bin → Gaussian 15

(d) Application of Bernoulli/Binomial to HAS-BLED 19

_____ Summary Section _____

Here is a quick statement of the 8 theorems:

Theorem 1: If X and Y are independent random variables, with p(x) and p(y) arbitrary, then

var(X+Y) = var(X) + var(Y).

Theorem 2: $\text{cov}(X,Y) = E(XY) - E(X)E(Y) = \Sigma_{x,y}\, xy\ p(x,y) - \mu_x\mu_y$.

Theorem 3: $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y)$.

Theorem 4: $\text{var}(aX+bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\,\text{cov}(X,Y)$

Theorem 5. $\text{var}(\Sigma_i a_i X_i) = \Sigma_{ij}\, a_i a_j\, \text{cov}(X_i,X_j)$

Corollary 5a. $\text{var}(\Sigma_i X_i) = \Sigma_{ij}\, \text{cov}(X_i,X_j)$ just setting all $a^i = 1$

Corollary 6a: $\text{var}(\Sigma_i X_i) = \Sigma_i\, \text{var}(X_i)$ if all X_i are independent

Theorem 7 (Standard Deviation of the Mean Theorem) For N repeats of the same experiment which have outcomes x (for random variable X), if the experiments are independent (they are) then $\sigma(\bar{x}) = \sigma/\sqrt{N}$ where $\sigma(\bar{x})$ is the standard deviation of the mean and σ is the standard deviation for one experiment. The pdf shape for the single experiment can be anything.

Theorem 8: (Bernoulli Trial SD of the Mean Theorem) . The standard deviation of the mean where the underlying process is a simple Bernoulli trial is given by $\sigma(\bar{x}) = \sqrt{pq/N}$.

Application: This theorem applies to "population studies" (including medical trials). Each "experiment" involves a different person (like a die throw), and the outcome might be how tall that person is, or whether or not that person had a major bleed over the period of a year.

For height, you assume for 7 billion people there exists some $\mu = \bar{x}$ and σ. You do an experiment with N people, just measuring height. You then plot all your heights in a histogram chart (they are all between 4' and 7', with mean 6', say). From this chart you come up with μ_1 and σ_1 -- these are your experiment's estimates for the true values μ and σ. Then your conclusion is that the average height is μ_1 = 6' and the standard deviation of mean is $\sigma(\bar{x}) = \sigma/\sqrt{N}$. Obviously large N reduces this standard deviation. The true mean μ for the full population lies within a few $\sigma(\bar{x})$ values of your μ_1 experimental estimate.

This conclusion is valid regardless of the nature of the underlying probability function which describes human height distribution. It could be a binomial distribution, a normal distribution, or any other kind of distribution, symmetric or asymmetric about the mean. The symbol "σ" means $\sqrt{\text{variance}}$, and every distribution, symmetric or not about the mean, has some variance $\text{var} = \Sigma_x\, p(x)\, (x-\mu)^2$.

_____ End of Summary Section _____

A. Review of Ng notes and Throwing one Die

First, from the wiki page http://en.wikipedia.org/wiki/Standard_error

SEM = standard error of the mean = $\sigma/\sqrt{n}$
σ = standard deviation (they call it s) "of the population" = $\sqrt{\text{var}}$

From my Ng notes we have,

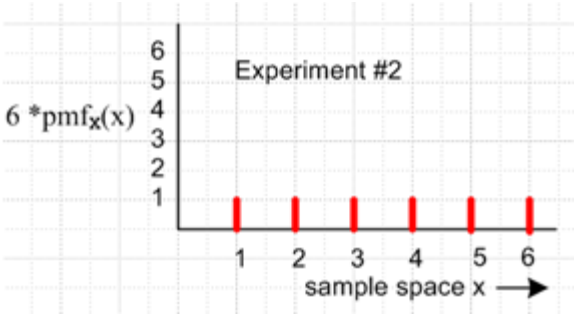
$$\text{var}(X) = \int_{-\infty}^{\infty} (x-\mu_x)^2\, p(x)\, dx. \qquad \mu_x = \int_{-\infty}^{\infty} x\, p(x)\, dx\ .$$

$$\text{var}(X_1) = (1/n)\, \Sigma_{i=1}^{n}\, (x_1{}^{(i)} - \mu_{X1})^2 \qquad \mu_{X1} = (1/n)\, \Sigma_{i=1}^{n}\, x^{(i)}{}_1 \qquad p = 1/n$$

$$\sigma_x^2 \equiv \text{var}_x = E(X^2) - \mu_x^2$$

In my Fourier Transform doc I have a huge appendix G on probability theory, hurray, I forgot all about it! I have just read it. I understand how to compute things for throwing 1 or 2 or N dice.

Dice Review. If you roll a fair die one time, you get this pmf,



For this distribution, you can compute whatever statistical object you want. For example,

$\mu = \Sigma_x\, x\, p(x) = \Sigma_x\, x\, (1/6) = (1/6)\, \Sigma_x x = (1/6)\, [\, 1 + 2 + 3 + 4 + 5 + 6] = (1/6)\, [21] = 7/2 = 3.5$

which seems pretty reasonable, see picture above. Another object:

$E(X^2) = \Sigma_x\, x^2\, p(x) = (1/6)\, \Sigma_x x^2 = (1/6)\, [\, 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2]$

$= (1/6)\, [\, 1 + 4 + 9 + 16 + 25 + 36] = (1/6)\, [\, 91] = 91/6 = [7*13/6] = 15.16666$

$\text{var} = E(X^2) - \mu_x^2 = 91/6 - (7/2)^2 = 15.16666 - (3.5)*(3.5) = 2.91666$

Well, in Appendix G I compute this by brute force enumerating all the terms and find that

$\text{var}(Z) = 5.833$

Notice that this is exactly twice the variance of the single die experiment!

$2* 2.91666 = 5.83332$

Why is this the case?? I will now answer that question with Theorem 1 below.

B. Theorems about Variance and Covariance

Consider rolling two such dice and examine z = x+y. We then have

$p(z) = \Sigma^z{}_{x,y}\, p(x,y)$ // App G

$E(Z) = \Sigma_z\, z\, p(z) = \Sigma_z\, z\, \Sigma^z{}_{x,y}\, p(x,y) = \Sigma_z\, \Sigma^z{}_{x,y}\, [zp(x,y)] = \Sigma_{x,y}\, (x+y)p(x,y)$

Here I have used my Lemma (G.15a). Next,

$E(Z^2) = \Sigma_z\, z^2\, p(z) = \Sigma_z\, z^2\, \Sigma^z{}_{x,y}\, p(x,y) = \Sigma_z\, \Sigma^z{}_{x,y}\, [z^2 p(x,y)] = \Sigma_{x,y}\, (x+y)^2 p(x,y)$

Therefore

$$\text{var}(Z) = E(Z^2) - [E(z)]^2 = \Sigma_{x,y}\,(x+y)^2 p(x,y) - [\,\Sigma_{x,y}\,(x+y)p(x,y)]^2$$

Now suppose we have $p(x,y) = p(x)p(y)$. Then

$$\text{var}(Z) = \Sigma_{x,y}\,(x+y)^2 p(x)p(y) - [\,\Sigma_{x,y}\,(x+y)p(x)p(y)]^2$$

$$= \Sigma_x \Sigma_y\,(x+y)^2 p(x)p(y) - [\,\Sigma_x \Sigma_y\,(x+y)p(x)p(y)]^2$$

$$= \Sigma_x p(x)\,\Sigma_y\,p(y)(x+y)^2 - [\,\Sigma_x p(x)\Sigma_y p(y)\,(x+y)]^2 = T1 - (T2)^2$$

$$T1 = \Sigma_x p(x)\,\Sigma_y\,p(y)(x^2 + 2xy + y^2)$$

$$= \Sigma_x p(x)\,\Sigma_y\,p(y)\;x^2 + \Sigma_x p(x)\,\Sigma_y\,p(y)\;2xy + \Sigma_x p(x)\,\Sigma_y\,p(y)\,y^2$$

$$= \Sigma_x p(x)\;x^2 + 2\,\Sigma_x p(x)x\,\Sigma_y\,p(y)\,y + \Sigma_y\,p(y)\,y^2$$

$$= E(X^2) + E(Y^2) + 2\,E(X)E(Y)$$

$$T2 = \Sigma_x p(x)\Sigma_y p(y)\,(x+y) = \Sigma_x p(x)\Sigma_y p(y)\,x + \Sigma_x p(x)\Sigma_y p(y)y$$

$$= \Sigma_x p(x)\,x + \Sigma_y p(y)y = E(X) + E(Y)$$

Therefore

$$\text{var}(Z) = T1 - (T2)^2 = E(X^2) + E(Y^2) + 2\,E(X)E(Y) - [E(X) + E(Y)]^2$$

$$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 = \text{var}(X) + \text{var}(Y)$$

So there is the theorem I have been looking for.

Theorems 1 through 6

Theorem 1: If X and Y are independent random variables, with p(x) and p(y) arbitrary, then

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y).$$

Back up and take a different pathway:

$$E(Z) = \Sigma_{x,y}\,(x+y)p(x,y)$$

$$E(Z^2) = \Sigma_{x,y}\,(x+y)^2 p(x,y)$$

Now recall that

$$\text{cov}(X,Y) = \Sigma_{x,y}\,(x-\mu_x)(y-\mu_y)\,p(x,y)$$

$= \Sigma_{x,y}\, xy\ p(x,y) - \mu_x \Sigma_{x,y} y\ p(x,y) - \mu_y \Sigma_{x,y}\ x\ p(x,y) + \mu_x\mu_y \Sigma_{x,y}\ p(x,y)$

$= \Sigma_{x,y}\, xy\ p(x,y) - \mu_x \Sigma_y\, y\, \Sigma_x\, p(x,y) - \mu_y\, \Sigma_x\, x\, \Sigma_y\, p(x,y) + \mu_x\mu_y$

$= \Sigma_{x,y}\, xy\ p(x,y) - \mu_x \Sigma_y\, y\, p(y) - \mu_y\, \Sigma_x\, x\, p(x) + \mu_x\mu_y$

$= \Sigma_{x,y}\, xy\ p(x,y) - \mu_x\mu_y - \mu_y\, \mu_x + \mu_x\mu_y$

$= \Sigma_{x,y}\, xy\ p(x,y) - \mu_x\mu_y$

$= E(XY) - E(X)E(Y)$

This is another new fact to me, so

Theorem 2: $\text{cov}(X,Y) = E(XY) - E(X)E(Y) = \Sigma_{x,y}\, xy\ p(x,y) - \mu_x\mu_y$.

Obviously if X and Y are independent, the result is 0, and if X = Y we get the usual result. But this Theorem 2 is a more general result that both of these claims.

Now back up to

$E(Z) = \Sigma_{x,y}\, (x+y)p(x,y)$

$E(Z^2) = \Sigma_{x,y}\, (x+y)^2 p(x,y)$

Then write

$E(Z^2) = E[(X+Y)^2] = E(X^2) + E(Y^2) + 2E(XY)$

$= E(X^2) + E(Y^2) + 2\ [\ \text{cov}(X,Y) + E(X)E(Y)\ ]$

$E(Z) = E(X+Y) = E(X) + E(Y)$

Then

$\text{var}(Z) = E(Z^2) - [E(Z)]^2 = E(X^2) + E(Y^2) + 2\ [\ \text{cov}(X,Y) + E(X)E(Y)\ ] - \{\ E(X) + E(Y)\}^2$

$= \text{var}(X) + \text{var}(Y) + 2\ [\ \text{cov}(X,Y) + E(X)E(Y)\ ] - 2\ E(X)E(Y)$

$= \text{var}(X) + \text{var}(Y) + 2\ \text{cov}(X,Y)$

So another general theorem:

Theorem 3: $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\ \text{cov}(X,Y)$.

Theorem 4: $\text{var}(aX+bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\ \text{cov}(X,Y)$

Now generalize for adding more terms. How would that go?

$Z = \Sigma_i\ a_i X_i$

$Z^2 = \Sigma_i\ a_i X_i\ \ \Sigma_j\ a_j X_j = \Sigma_{ij}\ a_i a_j X_i X_j$

$E(Z) = \Sigma_i\ a_i\ E(X_i)$

$E(Z^2) = \Sigma_{ij}\ a_i a_j\ E(X_i X_j)$

Then

$\text{var}(Z) = E(Z^2) - [E(Z)]^2 = \Sigma_{ij}\ a_i a_j\ E(X_i X_j) - [\ \Sigma_i\ a_i\ E(X_i)]^2$

Now write

$E(X_i X_j) = \text{cov}(X_i,X_j) + E(X_i)E(X_j)$

to get

$\text{var}(Z) = \Sigma_{ij}\ a_i a_j\ \{\ \text{cov}(X_i,X_j) + E(X_i)E(X_j)\} - [\ \Sigma_i\ a_i\ E(X_i)]^2$

$= \Sigma_{ij}\ a_i a_j\ \text{cov}(X_i,X_j) + \Sigma_{ij}\ a_i a_j E(X_i)E(X_j) - [\ \Sigma_i\ a_i\ E(X_i)]^2$

$= \Sigma_{ij}\ a_i a_j\ \text{cov}(X_i,X_j) + \Sigma_i a_i E(X_i) \Sigma_j a_j E(X_j) - [\ \Sigma_i\ a_i\ E(X_i)]^2$

$= \Sigma_{ij}\ a_i a_j\ \text{cov}(X_i,X_j) + [\Sigma_i a_i E(X_i)][\Sigma_j a_j E(X_j)] - [\ \Sigma_i\ a_i\ E(X_i)]^2$

$= \Sigma_{ij}\ a_i a_j\ \text{cov}(X_i,X_j)$

Suppose we had only two X_i. This would be

$\text{var}(a_1X_1 + a_2X_2) = a_1^2\ \text{cov}(X_1,X_1) + 2\ a_1a_2\ \text{cov}(X_1,X_2) + a_2^2\ \text{cov}(X_2,X_2)$

$= a_1^2\ \text{var}(X_1) + 2\ a_1a_2\ \text{cov}(X_1,X_2) + a_2^2\ \text{var}(X_2)$

and this reduces to Theorem 4. So we have an intermediate new theorem:

Theorem 5. $\text{var}(\Sigma_i a_i X_i) = \Sigma_{ij}\ a_i a_j\ \text{cov}(X_i,X_j)$

This also appears on the wiki variance page Good. Now break the sum into pieces

$$\Sigma_{ij}\, a_i a_j \,\mathrm{cov}(X_i,X_j) \;=\; \Sigma_{i=j}\, a_i a_j \,\mathrm{cov}(X_i,X_j) \;+ \Sigma_{i\neq j}\, a_i a_j \,\mathrm{cov}(X_i,X_j)$$

$$= \Sigma_i\, a_i^2 \,\mathrm{var}(X_i) \;+ \Sigma_{i\neq j}\, a_i a_j \,\mathrm{cov}(X_i,X_j)$$

$$= \Sigma_i\, a_i^2 \,\mathrm{var}(X_i) \;+ 2\,\Sigma_{i>j}\, a_i a_j \,\mathrm{cov}(X_i,X_j)$$

and this also appears on that same wiki variance page.

Corollary 5a. $\mathrm{var}(\Sigma_i X_i) \;=\; \Sigma_{ij}\,\mathrm{cov}(X_i,X_j)$ just setting all $a^i = 1$

Now suppose all the X_i are independent. Then we get

Theorem 6. $\mathrm{var}(\Sigma_i a_i X_i) \;=\; \Sigma_i\, a_i^2\,\mathrm{var}(X_i)$ if all X_i are independent.

Corollary 6a: $\mathrm{var}(\Sigma_i X_i) \;=\; \Sigma_i\,\mathrm{var}(X_i)$ if all X_i are independent

This was the theorem I was looking for when I started looking into things.

Example: Suppose you roll a die two times. Then

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

And if you roll it N times,

$$\mathrm{var}(X+Y+\ldots) = N\,\mathrm{var}(X)$$

Theorem 7 (Standard Deviation of the Mean Theorem) For N repeats of the same experiment which have outcomes x (for random variable X), if the experiments are independent (they are) then $\sigma(\bar{x}) = \sigma/\sqrt{N}$ where $\sigma(\bar{x})$ is the standard deviation of the mean and σ is the standard deviation for one experiment. The pdf shape for the single experiment can be anything.

Proof: Suppose you roll the dice N times (or any other experiment, X being some outcome). Then

$T \equiv X_1 + X_2 + \ldots.X_N$ = the function used in Corollary 6a.

so

$$\mathrm{var}(T) \;=\; \Sigma_i\,\mathrm{var}(X_i) = N\sigma^2$$

if all the processes are the same, like rolling a die N times. Now this can be restated as

$T/N = (X_1 + X_2 + \ldots.X_N)/N = (1/N)X_1 + \ldots.$ $a_i = 1/N.$

Then apply Theorem 6

$$\mathrm{var}(T/N) \;=\; \Sigma_i\,(1/N^2)\,\mathrm{var}(X_i) = (1/N^2)\Sigma_i\,\mathrm{var}(X_i) = (1/N^2)\,N\,\mathrm{var}(X)$$

$= \text{var}(X)/N \quad = \sigma^2/N$

Now T/N is the random variable for the mean value of all your samples. Call that $\bar{x}$. Then

$\sigma^2(\bar{x}) = \sigma^2/N$ = the variance of the mean value of your repeated X measurements.

or

$\sigma(\bar{x}) = \sigma/\sqrt{N}$

where σ is the SD of the "underlying experiment". QED.

This theorem and its proof are based on http://en.wikipedia.org/wiki/Standard_error .

Example: Suppose you roll a die N = 100 times. We know that $\sigma^2 = 2.91666$ from above, so we then know that σ = 1.708. The mean is μ = 3.5. Then

$\sigma(\bar{x}) = 1.708/\ 10 = .1708$

You could then add and subtract this to get your 1 std deviation range for the result.

Comment: Nothing has been said so far about any kind of "distribution". We have not specified anything to be a Gaussian distribution, for example. We assume that our random variable has SOME distribution, and therefore it has some variance and some σ. You still have variance and σ if the single event distribution is not symmetrical about its mean.

In the special case that you have a symmetric distribution for X_i, then I think T/N will also have a symmetric distribution. If the distribution is Gaussian, then you get 95% CI if you use 1.96 $\sigma(\bar{x})$ on either side.

"observation from a population" = observation of face-up number on a rolled die.

Comment: In HAS-BLED, the thing that corresponds to rolling the die is more like flipping a coin. For each of the N people in a HAS-BLED bin, there is a probability of a bleed which we call p.

I have now exhausted the wiki page on "standard error".

C. The Binomial Distribution and the Bernoulli Trial

Now I think the binomial distribution is going to get involved for HAS-BLED. For each person in the group of N = 500, either he bleeds or he doesn't bleed. Suppose p = probability of a person bleeding.

Note added: Flipping a weighted coin represents any Bernoulli Trial type process with A or B outcome. It is the repeated Bernoulli Trial which results in the binomial distribution.

Now digress to flipping N weighted coins where p = probability of a head. To get n heads, the probability is $(N,n)\ p^m q^{N-m}$. In other words

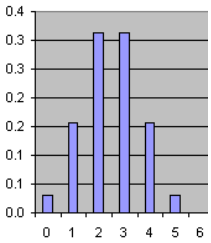
$$\text{prob}_N(\text{n heads}) = (N,n)\, p^n q^{N-n}$$

If p = 1/2, you get

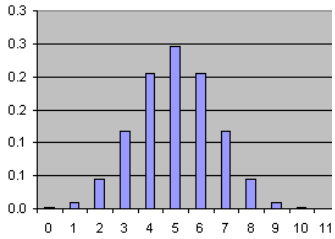
$$\text{prob}_N(\text{n heads}) = (N,n)\,(1/2)^n = [\, N!/(n!\,(N-n)!\,]\, 2^{-n}$$

The curve is symmetric if p = 1/2 because you can see that $\text{prob}_N(n) = \text{prob}_N(N-n)$.

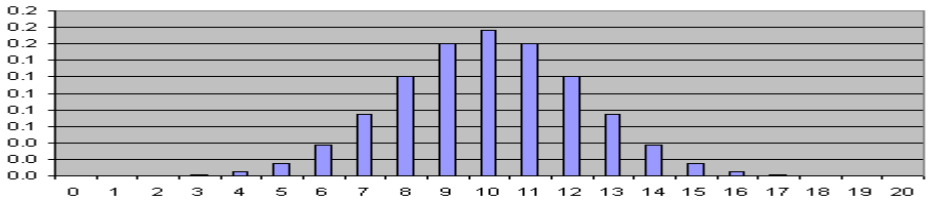
Here are some plots



N = 5

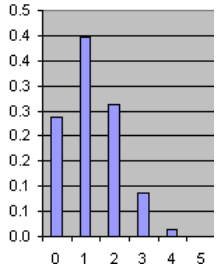


N = 10

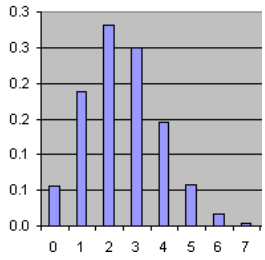


N = 20

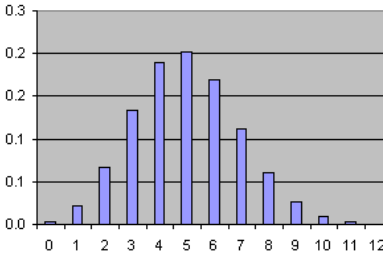
But if p = 0.25, things are different, the distribution is not symmetric:



N = 5



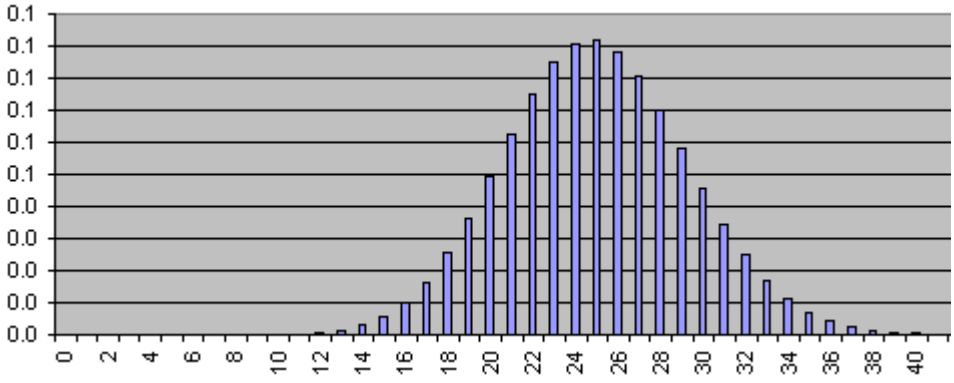
N = 10, p = 1/4



N = 20, p = 1/4

Now the distribution is NOT symmetrical. For small N it gets smashed into the left wall.

However, even if p ≠ q, for large N the distribution becomes close to symmetrical again.



N = 100, p=1/4

The mean value is Np = 100*1/4 = 25.

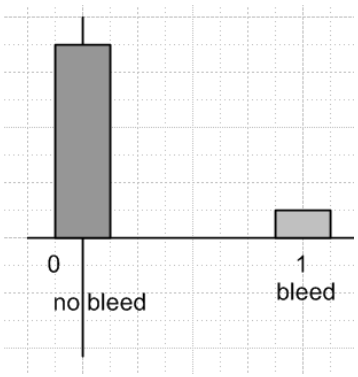
Here is some wiki stuff I could derive for this distribution, and we now rename N to be n, and n to be k

pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lfloor (n+1)p \rfloor - 1$
Variance	$np(1-p)$

Now the roll of the die here is the test of one patient for bleed or no bleed. Let's do this

no-bleed = 0
bleed = 1 // like a coin toss.

The PMF for this underlying experiment looks like this



The Bernoulli Trial Picture

What is σ for this thing? Well

var = $E(X^2) - [E(X)]^2$ E(X) = p, probability of a bleed

p(x) = p for x = 1 // this is the pmf
= q for x = 0

$[E(X)] = \Sigma_x \, x \, p(x) = 0*q + 1*p = p$

$E(X^2) = \Sigma_x \, x^2 \, p(x) = 0^2 *q + 1^2*p = p$

$var = p - p^2 = p(1-p) = pq = \sigma^2$

If you repeat this experiment (like coin toss) N time.

Now we want to know the standard deviation of the mean for doing this N times. The answer is

$\sigma(\bar{x}) = \sigma/\sqrt{N} = \sqrt{pq}/\sqrt{N} = \sqrt{pq/N}$

The mean here is $\bar{x} = E(x) = p$, as just shown above. So I have just proven:

Theorem 8: (Bernoulli Trial SD of the Mean Theorem) . The standard deviation of the mean where the underlying process is a simple Bernoulli trial is given by $\sigma(\bar{x}) = \sqrt{pq/N}$.

Proof. We know from Theorem 7 that in general $\sigma(\bar{x}) = \sigma/\sqrt{N}$ where σ is for the single experiment, For the Bernoulli experment, as shown above $\sigma = \sqrt{pq}$. QED.

Here is my older statement.

Theorem 8: (Bernoulli Trial SD of the Mean Theorem) For an experiment with a binary outcome with X = 1 or X = 0, we assume there is some pmf which can be characterized entirely by prob(1) = p. If this experiment is repeated N times, the mean value of the outcome will be p if N is a very large number. But for finite N, there is some distribution for the outcome (which is the binomial distribution) and we find that $\sigma(\bar{x}) = \sqrt{pq/N}$ where $\sigma(\bar{x})$ is the standard deviation of the mean p. In some sense, the mean you measure with finite N will lie in the distribution and the half width of the distribution in some sense is this $\sigma(\bar{x})$. Even for very small N, this is all true, but for such small N the binomial distribution is not symmetric about the mean, as my plots above show, so you have to think about how to construct something like a confidence interval for small N.

I first say the claim of this theorem on the wiki page for margin of error, where it appears this way:

$$\text{Standard error} \approx \sqrt{\frac{p(1-p)}{n}}$$

Go back now to

$$\text{prob}_N(\text{n heads}) = (N,n)\, p^n q^{N-n}$$

(a) I try but fail to show that Binomial Distribution approaches a Gaussian for large N.

In what sense does this approach a Gaussian for large N? I sort of see it in my graphs for p = 1/2. Stirling tells us that

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

So we can certainly claim that

$$N! \approx \sqrt{2\pi N}\ (N/e)^N$$

I suppose for some reasonable p value, everything of interest is large so then

$$\frac{N!}{n!\ (N-n)!} \approx \frac{\sqrt{2\pi N}\ (N/e)^N}{\sqrt{2\pi n}\ (n/e)^n \sqrt{2\pi (N-n)}\ ((N-n)/e)^{(N-n)}}$$

We end up with one factor of $\sqrt{2\pi}$ on the bottom. Power of e are these

$$e^{-N} * e^n * e^{N-n} = 1$$

so they all go away Then we have

$$N^N * n^{-n} * (N-n)^{-(N-n)}$$

I guess we now assume N >> n of interest so

$$\approx N^N * n^{-n} * (N)^{-(N-n)} = n^{-n} * (N)^{-(-n)} = n^{-n} N^n = (N/n)^n$$

This seems wrong, but continue. Then

$$\frac{N!}{n!\,(N-n)!} \approx (1/\sqrt{2\pi}) * (N/n)^n$$

and then

$$\text{prob}_N(\text{n heads}) \approx (1/\sqrt{2\pi}) * (N/n)^n p^n q^{N-n}$$

The mean value is $x_0 = Np$, and let $x = np$. Then $n = x/p$ so

$$\text{prob}_N(\text{n heads}) \approx (1/\sqrt{2\pi}) * (N/n)^n p^n q^{N-n}$$

$$= (1/\sqrt{2\pi}) * (Np/x)^{(x/p)} p^{(x/p)} q^{N-x/p}$$

$$= (1/\sqrt{2\pi}) * (x_0/x)^{(x/p)} p^{(x/p)} q^{N-x/p}$$

I don't see how this would ever have the form $\exp(- (x-x_0)^2/A)$. Keep going

$$= (1/\sqrt{2\pi}) * (x_0/x)^{(x/p)} p^{(x/p)} q^{(Np-x)/p}$$

$$= (1/\sqrt{2\pi}) * (x_0/x)^{(x/p)} p^{(x/p)} q^{(x0-x)/p}$$

$$= (1/\sqrt{2\pi}) * e^{(x/p)\ln(x0/x)} e^{\ln(x/p)} e^{\ln[(x0-x)/p]}$$

$$= (1/\sqrt{2\pi}) * e^{(x/p)[\ln(x0)-\ln(x)]} e^{\ln(x/p)} e^{\ln[(x0-x)/p]}$$

$$= (1/\sqrt{2\pi}) * e^{(x/p)[\ln(x0)-\ln(x)]} e^{[\ln x - \ln p]} e^{[\ln(x0-x)-\ln p]}$$

$$= (1/\sqrt{2\pi}) * e^{(x/p)[\ln(x0)-\ln(x)]} e^{\ln x} e^{\ln(x0-x)} e^{-2\ln p}$$

For large x the leading term dominates since it has an exposed x.

(b) Various Binomial Notations and Access from Excel or Maple

I need to find a source on this issue. What is b(n,p,j) ? Here is one notation from wiki on binomial,

$$f(k;n,p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

so yes, in my notation the variables are N, n and p so

$\text{prob}_{\mathbf{N}}$(n heads) = (N,n) $p^{\mathbf{n}}q^{\mathbf{N-n}}$ = f(n; N,p)

Here is another notation from my 12 chapter PDF book,

$$b(n,p,j) = \binom{n}{j} p^j q^{n-j}$$

and I would have

(N,n) $p^{\mathbf{n}}q^{\mathbf{N-n}}$ = b(N,p,n)

Excel has this

BINOMDIST(number_s,trials,probability_s,cumulative)
Returns the individual term binomial distribution probability.

-

BINOMDIST(number_s,trials,probability_s,cumulative)

Number_s is the number of successes in trials.

Trials is the number of independent trials.

Probability_s is the probability of success on each trial.

Cumulative is a logical value that determines the form of the function. If cumulative is TRUE, then BINOMDIST returns the cumulative distribution function, which is the probability that there are at most number_s successes; if FALSE, it returns the probability mass function, which is the probability that there are number_s successes.

So the last argument is just a flag which I would set false. Then I guess

BINOMDIST(n,N,p,FALSE) = b(N,p,n)

In maple we have

binomial(n r) = n!/r!/(n-r)!

which is just the coefficient part which Excel calls COMBIN(n,r). Maple doesn't have the full function, you have to make it yourself like so

b := (n,p,j) -> binomial(n j) p^j $(1-p)^{n-j}$

So, what does my big PDF book have to say on this?

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

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

μ = 0 and σ = 1 of the above

$$b(n, p, j) = \binom{n}{j} p^j q^{n-j}$$

Bernoulli Trials

Consider a Bernoulli trials process with probability p for success on each trial. Let $X_i = 1$ or 0 according as the ith outcome is a success or failure, and let $S_n = X_1 + X_2 + \cdots + X_n$. Then S_n is the number of successes in n trials. We know that S_n has as its distribution the binomial probabilities $b(n, p, j)$. In Section 3.2,

So we now have a term for what I am doing "Bernoulli Trials".

Definition 9.1 The *standardized sum* of S_n is given by

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \ .$$

S_n^* always has expected value 0 and variance 1.

Well OK this thing S_n^* is set up to have a mean of 0, AND it is scaled in a certain way.

(c) The Central Limit Theorem and my Second Failed Attempt to show Bin → Gaussian

Now here is the big connection claim:

Theorem 9.1 (Central Limit Theorem for Binomial Distributions) For the binomial distribution $b(n, p, j)$ we have

$$\lim_{n\to\infty} \sqrt{npq}\, b(n, p, \langle np + x\sqrt{npq}\rangle) = \phi(x) \ ,$$

where $\phi(x)$ is the standard normal density.

The author however does not prove this. I would write it out as

$\sqrt{npq}$ b(n,p,<x_0 + x$\sqrt{npq}$ >) → exp(-x^2/2) /$\sqrt{2\pi}$ as n→∞

Now back up a little and go back to

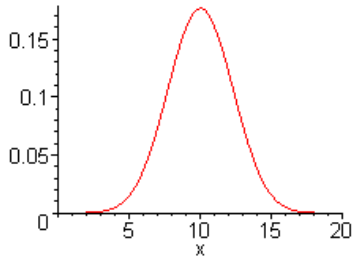
b := (n,p,j) -> binomial(n j) p^j (1-p)$^{n-j}$ b(n,p,j)

When you start, n and j are integers, but the gamma functions continue this thing to general values. So perhaps of interest is b(n,p,x) where we let x be a continuum

```
b := (n,p,x) ->binomial(n,x)*(p^x)* (1-p)^(n-x);
```

$$b := (n, p, x) \to \text{binomial}(n, x)\, p^x\, (1-p)^{(n-x)}$$

```
plot(b(20,0.5,x),x=0..20);
```



Now you know that the SD of this thing is σ = $\sqrt{npq}$ and the mean is μ = np. So the claim is

$\sqrt{npq}$ b(n,p,μ+xσ) → exp(-x^2/2) /$\sqrt{2\pi}$ as n→∞

Write this as

$\sqrt{npq}$ b(n,p,y) → exp(-x^2/2) /$\sqrt{2\pi}$ as n→∞ y = μ + σx

At x = 0, y = μ and the left side has it max. RHS also max at x = 0. Good.
At x = 1, y = μ+σ and we are one SD off the peak, more or less, and so is RHS.

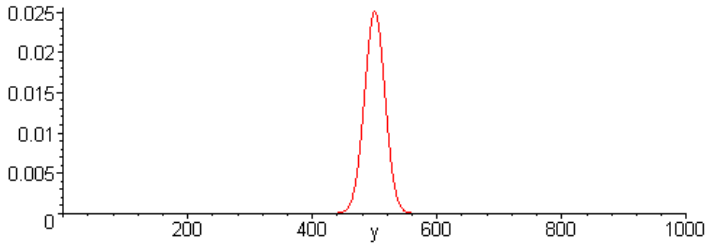
Let's try one more time on this.

LHS = σ b(n,p,y) = σ (n,y) p^y q^{n-y}

$$= \sigma \frac{n!}{y!\,(n-y)!}\, p^y\, q^{n-y}$$

Now look at our binomial plot for large n = 1000

```
plot(b(1000,0.5,y),y=0..1000);
```



In our range of non-zero-ness, we are going to have y large as well as n large. And n-y ≈ n/2 so it is also large. Then go ahead on all 3 with Stirling,

$$n! = \sqrt{2\pi n}\, n^n\, e^{-n}$$

$$\frac{n!}{y!\,(n-y)!} = \frac{\sqrt{2\pi n}\, n^n\, e^{-n}}{\sqrt{2\pi y}\, y^y\, e^{-y}\, \sqrt{2\pi(n-y)}\, (n-y)^{(n-y)}\, e^{-(n-y)}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{n^{n+.5}\, e^{-n}}{y^{y+.5}\, e^{-y}\, (n-y)^{(n-y)+.5}\, e^{-(n-y)}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{n^{n+.5}}{y^{y+.5}(n-y)^{(n-y)+.5}}$$

Then we have

$$\text{LHS} = \frac{1}{\sqrt{2\pi}} \frac{n^{n+.5}}{y^{y+.5}(n-y)^{(n-y)+.5}}\, \sigma\, p^y\, q^{n-y} \qquad y = \mu + \sigma x$$

Now the big question: what do you do next? All I can think to do is expo everything.

$$\text{num} = e^{(n+1/2)\ln(n)}\ \sigma\ e^{y\ln(p)}\ e^{(n-y)\ln(q)}$$

$$\text{den} = \sqrt{2\pi}\ e^{(y+1/2)\ln(y)}\ e^{(n-y+1/2)\ln(n-y)}$$

Then can combine into a single exp:

$$\text{LHS} = \frac{\sigma}{\sqrt{2\pi}} \exp\{\, (n+1/2)\ln(n) + y\ln(p) + (n-y)\ln(q) - (y+1/2)\ln(y) - (n-y+1/2)\ln(n-y)\,\}$$

If n, y and n-y are all large, as I have assumed, then the 1/2 contributions should not matter, then

$$\text{LHS} = \frac{\sigma}{\sqrt{2\pi}} \exp\{\, n \ln(n) + y\ln(p) + (n-y)\ln(q) - y \ln(y) - (n-y) \ln(n-y) \,\}$$

As before, I am completely stumped, no idea what to do next. How get this to be $\exp(-x^2/2)/\sqrt{2\pi}$?

It does seem that the terms like y ln(p) and (n-y)ln(q) are smaller than the other terms, so try

$$\text{LHS} \approx \frac{\sigma}{\sqrt{2\pi}} \exp\{\, n \ln(n) - y \ln(y) - (n-y) \ln(n-y) \,\} \qquad y = \mu + \sigma x$$

and at least now things don't depend on p and q. Maybe regroup terms,

$$\text{LHS} \approx \frac{\sigma}{\sqrt{2\pi}} \exp\{\, n \ln(n) - n \ln(n-y) - y \ln(y) + y \ln(n-y) \,\}$$

$$\text{LHS} \approx \frac{\sigma}{\sqrt{2\pi}} \exp\{\, n [\, \ln(n) - \ln(n-y)] - y [\, \ln(y) - \ln(n-y)] \,\}$$

$$\text{LHS} \approx \frac{\sigma}{\sqrt{2\pi}} \exp\{\, n [\, \ln(n) - \ln(n-y)] - y [\, \ln(y) - \ln(n-y)] \,\}$$

$$\ln(n) - \ln(n-y) = \ln \frac{n}{n-y} = -\ln \frac{n-y}{n} = -\ln(1 - y/n)$$

$$\ln(y) - \ln(n-y) = \ln \frac{y}{n-y} = -\ln \frac{n-y}{y} = -\ln(n/y-1)$$

Then we have

$$\text{LHS} \approx \frac{\sigma}{\sqrt{2\pi}} \exp\{\, n [\, -\ln(1 - y/n)] - y [\, -\ln(n/y-1] \,\}$$

$$\text{LHS} \approx \frac{\sigma}{\sqrt{2\pi}} \exp\{\, -n \ln(1 - y/n)] + y \ln(n/y-1] \,\}$$

Once again, I hit a dead end. I will need to find a proof.

A Bernoulli random variable $\text{Ber}(p)$ is 1 with probability p and 0 otherwise. A binomial random variable $\text{Bin}(n, p)$ is the sum of n independent $\text{Ber}(p)$ variables.

The Central Limit Theorem says that you end up with a normal distribution when you add a bunch of random variables no matter what you had for p(x) for that variable! So it is a very general theorem, and proofs I am looking at are rather involved affairs.

Wiki has a "simple proof".

(d) Application of Bernoulli/Binomial to HAS-BLED

Status: For the HAS-BLED study we have this

Patients With Particular Score Among Those Taking Warfarin Only*	Major Bleedin Events†
746 (20.4)	7 (0.9)
1,283 (35.0)	44 (3.4)
950 (25.9)	39 (4.1)
483 (13.2)	28 (5.8)
180 (4.9)	16 (8.9)
22 (0.6)	2 (9.1)
1 (0.0)	0

The numbers are the left are number of people in each bin. These numbers are all > 20 except the last one, but I think it would be just find to do 1.96 SD and just assume normal distribution for each. The key idea is this

$$\text{Standard error} \approx \sqrt{\frac{p(1-p)}{n}}$$

For example, try applying this to the bin with 22 people in it. We have

p = .091 rate of bleeds

n = 22

```
restart;
se := sqrt(p*(1-p)/n);
```

$$se := \sqrt{\frac{p\,(1-p)}{n}}$$

```
p := .091;
```

$$p := .091$$

```
n := 22;
```

$$n := 22$$

```
se;
```

$$.06131846822$$

```
1.96*se;
```

$$.1201841977$$

Recall,

$\sigma(\bar{x}) = \sigma/\sqrt{N} = \sqrt{pq}/\sqrt{N} = \sqrt{pq/N}$

My conclusion is that

rate = .091 ± .120 ≈ 0 to .21

Then the measured rate is 9.1%, and the actual rate with 95% confidence is in the range 0% to 21%.

Now let's try the next bin up:

```
p := .089;
```

$$p := .089$$

```
n := 180;
```

$$n := 180$$

```
se;
```

$$.02122354563$$

```
1.96*se;
```

$$.04159814943$$

Then the rate is .089 ± .042 = (.047, ,131) = 4.7% to 13.1%

I am now ready to do this in Excel. It took me all day to get ready?