Lecture 19: More expectations. Summing independent random variables; Fri Nov 2 (a) L & M 3.9. (Here we show the case of two discrete random variables, but the results are true in general.)

19.1 Expectation of a sum

Let X and Y be two (discrete) random variables with joint p.m.f

$$p_{X,Y}(x,y) \;=\; P(X=x \cap Y=y) \text{ for } x \in \mathcal{X}, y \in \mathcal{Y}$$

Then in 13.2 (ii) we showed $\mathrm{E}(g_1(X)+g_2(X)) \;=\; \mathrm{E}(g_1(X)+\mathrm{E}(g_2(X))$.

Now we just replace the single variable X with the pair (X,Y):

$$\mathrm{E}(g_1(X,Y)+g_2(X,Y)) \;=\; \mathrm{E}(g_1(X,Y)+\mathrm{E}(g_2(X,Y)).$$

In particular, with $X=g_1(X,Y)$ and $Y=g_2(X,Y)$, we have $\mathrm{E}(X+Y) \;=\; \mathrm{E}(X)+\mathrm{E}(Y)$.

Expectation of sum is sum of expectations: always; independence is **not** required.

19.2 Expectation of the product: case of independent random vaiables

In general: $\mathrm{E}(XY) \;=\; \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} x\ y\ P(X=x \ \cap\ Y=y) \;=\; \sum_x\sum_y\ x\ y\ p_{X,Y}(x,y)$.

If X and Y are **independent** then

$$\begin{aligned}\mathrm{E}(XY) &= \sum_x\sum_y x\ y\ p_{X,Y}(x,y) \;=\; \sum_x\sum_y x\ y\ P(X=x)P(Y=y)\\ &= \left(\sum_x x\ P(X=x)\right)\left(\sum_y y\ P(Y=y)\right) \;=\; \mathrm{E}(X)\mathrm{E}(Y).\end{aligned}$$

Note: The converse is NOT true. We can have $\mathrm{E}(XY) \;=\; \mathrm{E}(X).\mathrm{E}(Y)$, but X and Y **not** independent.

19.3 Covariance

Definition: $\mathrm{Cov}(X,Y) \;=\; \mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))$ Note $\mathrm{Cov}(X,X)=\mathrm{var}(X)$.

Note: Let $\mu=\mathrm{E}(X)$, $\nu=\mathrm{E}(Y)$, then

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}((X-\mu)(Y-\nu)) \;=\; \mathrm{E}(XY-\mu Y-\nu X+\mu\nu)\\ &= \mathrm{E}(XY)-\mu\mathrm{E}(Y)-\nu\mathrm{E}(X)+\mu\nu \;=\; \mathrm{E}(XY)-\mu\nu \;=\; \mathrm{E}(XY)-\mathrm{E}(X).\mathrm{E}(Y)\end{aligned}$$

So, if X and Y are independent, $Cov(X,Y)=0$. (Again, the converse is **not** true.)

19.4 Variance of the sum: general and then independence case

$$\begin{aligned}\mathrm{var}(X+Y) &= \mathrm{E}((X+Y)^2) \;-\; (\mathrm{E}(X+Y))^2\\ \mathrm{E}((X+Y)^2) &= \mathrm{E}(X^2+2XY+Y^2) \;=\; \mathrm{E}(X^2)+2\mathrm{E}(XY)+\mathrm{E}(Y^2)\\ (\mathrm{E}(X+Y))^2 &= (\mathrm{E}(X)+\mathrm{E}(Y))^2 \;=\; (\mathrm{E}(X))^2+2\ \mathrm{E}(X)\ \mathrm{E}(Y)+(\mathrm{E}(Y))^2\\ \mathrm{var}(X+Y) &= \mathrm{var}(X)\;+\;2\ \mathrm{Cov}(X,Y)\;+\;\mathrm{var}(Y)\end{aligned}$$

If X and Y are **independent**, $\mathrm{Cov}(X,Y) \;=\; 0$. Then: $\mathrm{var}(X+Y) \;=\; \mathrm{var}(X) \;+\; \mathrm{var}(Y)$,

For independent random variables, the variance of the sum is the sum of the variances.

Note: If we can sum 2, we can sum any finite number: $X+Y+Z \;=\; (X+Y) \;+\; Z$.

19.5 Examples See Binomial and other examples: 21.1 and then (after # **22**), 21.5 and 21.6.

19.6 Correlation

Definition: $\rho(X,Y) \;=\; \mathrm{Cov}(X,Y)/\sqrt{\mathrm{var}(X).\mathrm{var}(Y)}$. Note $\rho(X,X)=1$.

Result:

Let X have variance σ_X^2 and Y have variance σ_Y^2.

$$0 \;\le\; \mathrm{var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) \;=\; \frac{\mathrm{var}(X)}{\sigma_X^2} + \frac{\mathrm{var}(Y)}{\sigma_Y^2} \pm 2\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} \;=\; 2(1 \pm \rho(X,Y))$$

Hence $0\le(1-\rho(X,Y))$ so $\rho\le 1$; $0\le(1+\rho(X,Y))$ so $\rho\ge-1$. i.e. $-1\le\rho\le 1$.

Lecture 20: The Normal distribution: L & M 4.3; Fri Nov 2 (b)

20.1 The standard Normal probability density

A random variable Z with p.d.f $f_Z(z) \;=\; \frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})$ on $-\infty < z < \infty$ is a *standard Normal random variable*. Its pdf is often denoted $\phi(z)$, and its cdf $P(Z \leq z) \;=\; \Phi(z)$.

$$\int_{-\infty}^{\infty} f_Z(z)dz \;=\; 1 \qquad \text{Can be shown using polar coordinates, see 17.4}$$

$$\mathrm{E}(Z) \;=\; \int_{-\infty}^{\infty} z f_Z(z)dz \;=\; 0 \qquad \text{since } f_Z(-z) = f_Z(z).$$

$$\mathrm{var}(Z) \;=\; \mathrm{E}(Z^2) \;=\; \int_{-\infty}^{\infty} z^2 f_Z(z)dz \;=\; \int_{-\infty}^{\infty} f_Z(z)dz \;=\; 1 \quad \text{Integrate by parts: see 17.4}$$

20.2 The Normal probability density, parameters μ and σ^2: $N(\mu, \sigma^2)$

$$f_X(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \qquad -\infty < x < \infty$$

$$P(X \in B_x) \;=\; \int_{B_x} \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})\;dx$$

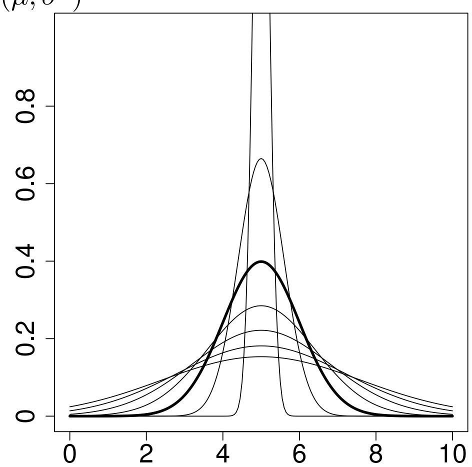
Let $z \;=\; (x-\mu)/\sigma$, $dz = dx/\sigma$, $B_z = \{(x-\mu)/\sigma; x \in B_x\}$

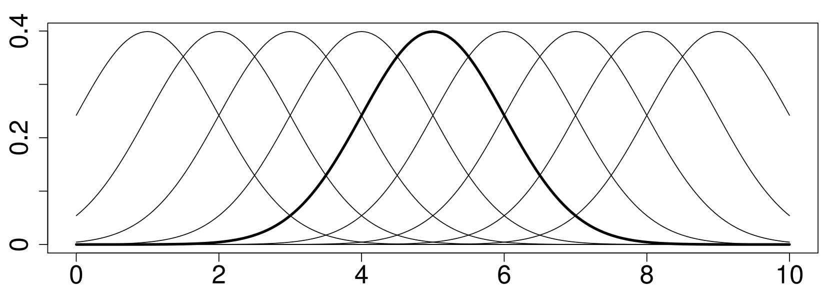
$$P(Z = (X-\mu)/\sigma \in B_z) \;=\; \int_{B_z} \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{z^2}{2})\;\sigma\;dz$$

$$= \int_{B_z} \frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})\;dz$$

That is $f_Z(z)$ is a Normal probability density with parameters 0 and 1.

Also, μ is a location parameter, σ is a scale parameter.





Above: Normal pdfs at different scales (values of σ).

Left: Normal pdfs at different locations (values of μ).

Now $Z = (X-\mu)/\sigma$ or $X = \mu + \sigma Z$. From 20.1, E(Z)=0 amd var(Z)=1, so $\mathrm{E}(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$.

20.3: Using the Normal probability table (Or, use the R-function *pnorm()*)

Table A1 in L & M (Pp 697-698) is of the usual form for the probabilities for a N(0,1) standard Normal distribution. It gives $\Phi(z) = P(Z \leq x)$ for values of x from -3 to $+3$.

In fact we only need half the table since the distribution is symmetric about 0.

For negative x, $P(Z \leq x) \;=\; P(Z \geq -x) \;=\; 1 - P(Z \leq -x)$.

(Note that since Z is a continuous random variable, $P(Z < x) \;=\; P(Z \leq x)$).

20.4: The Central Limit Theorem: statment only

Suppose, $Y_1, ..., Y_n$ are independent, with the same distribution, each with (finite) mean μ and (finite) variance σ^2.

If $T_n = \sum_{i=1}^{n} Y_i$, then $\mathrm{E}(T_n) = n\mu$ and $\mathrm{var}(T_n) = n\sigma^2$.

So $T_n^* \;=\; (T_n - n\mu)/(\sqrt{n}\;\sigma)$ has mean 0 and variance 1.

Subject to some conditions, for large n, T_n^* has approx. a $N(0,1)$ pdf: This is the **Central Limit Theorem**.

Note also, if $\overline{Y} = (T_n/n)$, then $T_n^* = \sqrt{n}(\overline{Y} - \mu)/\sigma$. (Giving the form for averages of r.vs, rather tha sums.)

Lecture 21: Normal approximation to the Binomial; L& M 4.3; Fri Nov 2(b) ctd.

21.1: The Binomial Mean and variance

(i) Let Y_i be independent Bernoullia random variables: where $P(Y_i = 1) = p$, $P(Y_i = 0) = 1 - p$.

$$\mathrm{E}(Y_i) = p,\ \mathrm{E}(Y_i^2) = p, \text{ so } \mathrm{var}(Y_i) = p - p^2 = p(1-p).$$

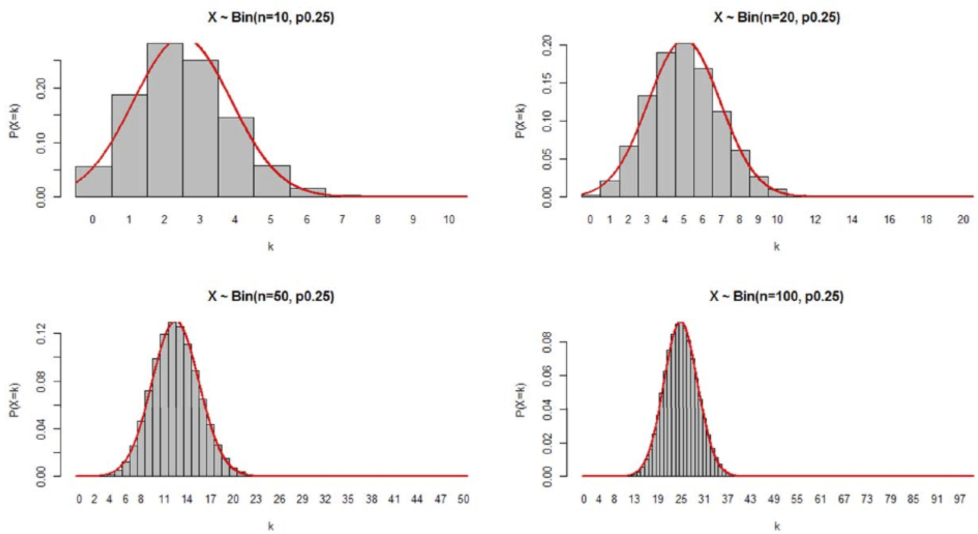
(ii) Then if $X \sim B(n,p)$, $X = \sum_{i=1}^{n} Y_i$, so $\mathrm{E}(X) = \sum_1^n \mathrm{E}(Y_i) = np$, and because the Y_i are independent, $\mathrm{var}(X) = \sum_1^n \mathrm{var}(Y_i) = np(1-p)$.

(iii) Applying the CLT, with $\mathrm{E}(Y_i) = \mu = p$, $\mathrm{var}(Y_i) = \sigma^2 = p(1-p)$,

$$(X - np)/\sqrt{np(1-p)} = (\sum_{i=1}^{n} Y_i - np)/\sqrt{np(1-p)}$$

is approximately $N(0,1)$ for large n.

21.2: Binomials vs Normals: examples



21.3: The continuity correction

(a) When approximating $P(X = k)$ for a binomial X by a Normal Y, we must consider

$P(k - \frac{1}{2} < Y \le k + 1/2)$. We need area under the curve.

(b) Hence when we approximate $P(a \le X \le b)$ we should use $P(a - \frac{1}{2} < Y < b + \frac{1}{2})$.

(c) However, for large n and a reasonable range of Y it makes almost no difference. Recall when X is increased by 1, Z is increased by $\delta = 1/\sqrt{np(1-p)}$.

21.4 Approximating discrete Binomial with continuous Normal

Example: suppose X is $Bin(30, 2/3)$. $\mathrm{E}(X) = 20$, $\mathrm{var}(X) = 30 \times (1/3) \times (2/3) = 20/3$.

Compute the probability $14 \le X \le 18$

(i) Exactly, using the Binomial probabilities: $\sum_{k=14}^{18} P(X = k)$. Answer 0.2689.

(ii) Using the Normal approx, with the range 14 to 18 for X:

$Z = (14 - 20)/\sqrt{20/3} = -2.32$ to $Z = (18 - 20)/\sqrt{20/3} = -0.77$. Answer: 0.2105.

(iii) Using the Normal approx, with the range 13.5 to 18.5 for X:

$Z = (13.5 - 20)/\sqrt{20/3} = -2.517$ to $Z = (18.5 - 20)/\sqrt{20/3} = -0.5809$. Answer: 0.2747.

For general a, b: $P(a < Z \le b) = \Phi(b) - \Phi(a)$.

22: Some additional notes about CLT for Binomials

22.1: Stirling's formula (Skip this in 2018)

For large n, $n!$ is approximately $n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$. Let $k = np$, then

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} = \frac{n!}{(np)!(n(1-p))!} \\
&\approx \frac{n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}}{(np)^{np+\frac{1}{2}}e^{-np}\sqrt{2\pi}(n(1-p))^{n(1-p)+\frac{1}{2}}e^{-n(1-p)}\sqrt{2\pi}} \\
&= \frac{n^{n+\frac{1}{2}}}{(np)^{np+\frac{1}{2}}(n(1-p))^{n(1-p)+\frac{1}{2}}\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{np(1-p)}}\frac{1}{p^{np}(1-p)^{n(1-p)}}
\end{aligned}$$

22.2: The DeMoivre-Laplace limit theorem (Central limit theorem for Binomials) **Skip for 2018**

(a) For a $Bin(n,p)$ random variable X, the p.m.f. is largest at $k \approx np$: $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$.

So for n large and $k = np$ we have

$$P(X = np) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{np(1-p)}}\frac{1}{p^{np}(1-p)^{n(1-p)}}\; p^{np}\;(1-p)^{n(1-p)} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{np(1-p)}}$$

But this is also the value of the maximum pdf of a Normal random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.

(b) As X increases from k to $k+1$, $Z = (X-np)/\sqrt{np(1-p)}$ increases from z to $z+\delta$ where $z = (k-np)/\sqrt{np(1-p)}$ and $\delta = 1/\sqrt{np(1-p)}$. (Note $np(1-p) = 1/\delta^2$.) For a Normal $N(np, 1/\delta^2)$ pdf;

$$\frac{f_Y(k+1)}{f_Y(k)} = \frac{\frac{\delta}{\sqrt{2\pi}}\exp(-\delta^2(k-np+1)^2/2)}{\frac{\delta}{\sqrt{2\pi}}\exp(-\delta^2(k-np)^2/2)} = \frac{\exp(-\delta^2((z/\delta)+1)^2/2)}{\exp(-\delta^2(z/\delta)^2/2)} = \exp(-z\delta - \delta^2/2) \approx (1-\delta z)$$

(c) But

$$\frac{P(X = (k+1))}{P(X = k)} = \frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k}p^k(1-p)^{n-k}} = \left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right)$$

$$\begin{aligned}
&= \left(\frac{n(1-p) - z/\delta}{np + (z/\delta) + 1}\right)\left(\frac{p}{1-p}\right) = \frac{np(1-p) - zp/\delta}{np(1-p) + (1-p)z/\delta \; + \; (1-p)} \approx \frac{1-\delta zp}{1+\delta z(1-p)} \\
&\approx (1-\delta zp)(1-\delta z(1-p)) = (1-\delta zp - \delta z(1-p)) = (1-\delta z)
\end{aligned}$$

22.3: Mendel's experiments *(Rest of* **#22** *mainly just for interest, but lots examples here!!)*

Mendel did many experiments of the form of the one with the red/white flowers. He crossed red-flowered plants with white-flowered plants, so he knew the red-flowered offspring were of RW type. These are known as the F_1 or *hybrids*. He then crossed these with each other, and expected to get red and white flowers in the ratio 3:1. Here are four examples:

a) 253 F_1 producing 7324 seeds: 5474 round, 1850 wrinkled: ratio 2.96:1

b) 258 F_1 producing 8023 seeds: 6022 yellow, 2001 green: ratio 3.01:1.

c) 929 F_2; 705 red flowers, 224 white flowers: ratio 3.15:1.

d) 580 F_2: 428 green pods, 152 yellow pods: ratio 2.82:1

22.4 Are Mendel's results too good?

There has been much debate as to whether Mendel's results are "too good" – too close to the 3:1 ratio.

Note the larger samples for characteristics that can be observed at the seed stage. These give the ratios closest to 3:1. This is as expected: $\mathrm{var}(X) = np(1-p)$ but $\mathrm{var}(X/n) = \mathrm{var}(X)/n^2 = p(1-p)/n$ which decreases as n increases. Are we too close? Recall $Z = (X - np)/\sqrt{np(1-p)}$ is approx N(0,1). Here $p = 3/4$:

a) $Z_a = (5474 - 7324 \times 0.75)/\sqrt{7324 \times 3/16} = -0.5127$, $P(-0.5127 < Z \leq 0.5127) = 2\Phi(0.5127) - 1 = 0.39$.

b) $Z_b = (6022 - 8023 \times 0.75)/\sqrt{8023 \times 3/16} = 0.1225$, $P(-0.1225 < Z \leq 0.1225) = 2\Phi(0.1225) - 1 = 0.097$.

c) $Z_c = (705 - 929 \times 0.75)/\sqrt{929 \times 3/16} = 0.6251$, $P(-0.6251 < Z \leq 0.06251) = 2\Phi(0.6251) - 1 = 0.468$.

d) $Z_d = (428 - 580 \times 0.75)/\sqrt{580 \times 3/16} = -0.6712$, $P(-0.6712 < Z \leq 0.6712) = 2\Phi(0.6712) - 1 = 0.498$.

So far, with these experiments, there seems no reason to think Mendel's results are "too good".

22.5 Combining the experiments

The fact that these involve different characteristics does not stop us combining them. They are all independent Bernoulli trials with $p = 0.75$.

We have $7324 + 8023 + 929 + 580 = 16856$ trials with $5474 + 6022 + 705 + 428 = 12629$ "successes". $Z = (12629 - 16856 \times 0.75)/\sqrt{16856 * 3/16} = -0.2312$. $P(-0.2312 < Z \leq 0.2312) = 2\Phi(0.2312) - 1 = 0.183$.

Alternatively, we can combine the Z-values: we could do this even if they came from Bernoulli trials with different p. Here: $Z_a + Z_b + Z_c + Z_d = -0.5127 + 0.1225 + 0.6251 - 0.6712 = -0.4363$.

This would be a Normal with mean 0 but variance 4 (why?). So we must standardize it:

$Z^* = -0.4363/2 = -0.2182$, $P(-0.2182 < Z \leq 0.2182) = 2\Phi(0.2182) - 1 = 0.173$.

So again, either way, here there is no evidence of the results being "too good". However, when a large number of Mendel's other results are also grouped together, overall, they do look a bit "too good".

22.6 Mendel's experiment: continued

Now Mendel wanted to show not just the 3:1 red:white ratio, but also the 1:2:1 for $RR : RW : WW$. So he needed to find which of his red-flowered F_2 plants were RR and which were RW. To do this he *selfed* his red-flowered F_2 pea plants: that is, the parents were RR giving $RR \times RR$ or RW giving $RW \times RW$.

In order to tell whether the parent was RW, Mendel grew up 10 offspring, and if all were red he said the plant *bred true*. Note, under Mendel's hypothesis $P(RR \mid \text{red}) = 1/3$.

Mendel reported his result: from 600 F_2 he found 201 *bred true*. Assuming 1/3 should *breed true*, is this result too close to 1/3? Note if $p = 1/3$, $\mathrm{E}(X) = 200$, $\mathrm{var}(X) = 600 \times 1/3 \times 2/3 = 400/3$.

(i) Without the correction (considering $X = 199, 200, 201$) show the probability of being this close is about 6.5%. ($Z = \pm 0.08660$).

(ii) With the correction ($198.5 < X < 201.5$) show the probability of being this close is a bit over 10% ($Z = \pm 0.12990$).

(Here the continuity correction makes enough difference that is might affect our belief about whether Mendel's results are "too good").

22.7 Mendel's mistake:

Recall that each offspring of an $RW \times RW$ mating is white with probability 1/4.

(i) For each $RW \times RW$ mating, what is the probability Mendel mis-called it as $RR \times RR$?

Answer: $(3/4)^{10} = 0.0563$.

(ii) If the frequency of RR parents is 1/3 and RW is 2/3, what is the overall probability that all 10 offspring plants are red? Answer: $(1/3) + (2/3) \times 0.0563 \ = \ 0.371$.

22.8 Probability of being close to 0.371

So now the p of Mendel's Binomial should have been $p = 0.371$. $\mathrm{E}(X) = 222.6$, $\mathrm{var}(X) = 140.01$, st.dev $=$ 11.83. Now we need the probability that Mendel's reported count of 201 would be *this far off*.

(i) With no correction: $X \leq 201$, $Z < -1.825$ or $Z > 1.825$. Answer: about 6.8%.

(ii) With correction: $X \leq 201.5$, $Z < -1.783$ or $Z > 1.783$. Answer: about 7.4%.

(iii) Or maybe we should ask, this far off in direction of his assumed 1/3, Asnwers: 3.4% and 3.7%.

Either Mendel was, for once, quite *unlucky* or else his result is too close to what he may have expected, and too far from what he should have found.

Back to section 21 and variances and covariances

21.5 Multinomial means, variances, and covariances

(i) Sample size n, k possible types, $P(\text{type } j) = p_j$. $\sum_{j=1}^{k} p_j \ = \ 1$. Let $Y_ij = 1$ if i^{th} is type j, and 0 otherwise. X_j $(j = 1, ..., k)$ is number of type j. $(X_1, ..., X_k)$ is multinomial. $\sum_{j=1}^{k} X_j \ = \ n$.

(ii) Note $X_j = \sum_{i=1}^{n} Y_{ij}$. But Y_{ij} is Bernoulli, mean p_j variance $p_j(1 - p_j)$.

Also the Y_{ij} are independent *over* i (NOT over j).

So $\mathrm{E}(X_j) = np_j$, $\mathrm{var}(X_j) = np_j(1 - p_j)$ – nothing new here $X_j \ \sim \ Bin(n, p_j)$.

(iii) Note $Y_{ij}Y_{i\ell} \equiv 0$: if i^{th} is type j is is not type ℓ.

So $\mathrm{Cov}(Y_{ij}, \ Y_{i\ell} \ = \ 0 - p_j p_\ell \ = \ -p_j p_\ell$. So

$$\begin{aligned}
\mathrm{Cov}(X_j, \ X_\ell) \ &- \ \mathrm{Cov}(\sum_i Y_{ij}, \ \sum_{i'} Y_{i'\ell}) \ = \ \sum_i \sum_{i'} \mathrm{Cov}(Y_{ij}, \ Y_{i'\ell}) \\
&= \ \sum_i \mathrm{Cov}(Y_{ij}, \ Y_{i\ell}) \ = \ \sum_i (-p_j p_\ell) \ = \ -np_j p_\ell
\end{aligned}$$

21.6 Hypergeometric mean and variance

(i) Sampling n items without replacement, from total N of which k are "red".

X is the number that are red: values $x = \max(0, n - (N - k)), ..., \min(n, k)$.

Let $Y_i = 1$ if i^{th} is red, and 0 otherwise. $P(Y_i = 1) = k/N = p$. $\mathrm{E}(Y_i) = p$ and $\mathrm{var}(Y_i) = p(1 - p)$.

Then $X \ = \ \sum_{i=1}^{n} Y_i$, and $\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(Y_i) \ = \ nk/N \ = \ np$. (Like Binomial).

(ii) $\mathrm{E}(Y_i Y_j) \ = \ P(\{Y_i = 1\} \bigcap \{Y_j = 1\}) \ = \ P(Y_i = 1 \mid Y_j \ = \ 1) \ P(Y_j \ = \ 1) = ((k-1)/(N-1)).(k/N)$.

So $\mathrm{Cov}(Y_i, \ Y_j) \ = \ (k(k-1))/(N(N-1)) \ - \ (k/N)^2 \ = \ -k(N-k)/(N^2(N-1)) \ = \ -p(1-p)/(N-1)$

(iii) Then

$$\begin{aligned}
\mathrm{var}(X) \ &= \ \mathrm{var}(\sum_{i=1}^{n} Y_i) \ = \ \sum_{i=1}^{n} \mathrm{var}(Y_i) \ + \ 2 \sum_{(i,j):} \sum_{i<j} \mathrm{Cov}(Y_i, Y_j) \\
&= \ np(1-p) + 2\binom{n}{2} \frac{(-p(1-p))}{(N-1)} \ = \ np(1-p)\frac{(N-n)}{(N-1)}
\end{aligned}$$

Smaller than corresponding Binomial variance.