

MATH 2030 3.00MW – Elementary Probability
Course Notes
Part IV: Binomial/Normal distributions
Mean and Variance

Tom Salisbury – salt@yorku.ca

York University, Dept. of Mathematics and Statistics
Original version April 2010. Thanks are due to
E. Brettler, V. Michkine, and R. Shieh for many corrections

January 18, 2018

Binomial Distribution

- ▶ The course now swings towards studying specific distributions and their applications. Along the way we'll define and study means and variances.
- ▶ Recall that if X has a binomial distribution $X \sim \text{Bin}(n, p)$ then $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, $k = 0, 1, \dots, n$. Here $0 \leq p \leq 1$ and n is a positive integer.
- ▶ We saw earlier (as an application of the binomial theorem) that these probabilities sum to 1, so this really is a dist'n.
- ▶ It arises from counting the number of successes in n repeated independent trials of some experiment. Each trial results in success or failure. We need that:
 - ▶ The trials are independent;
 - ▶ There is the same probability p of success in each trial.

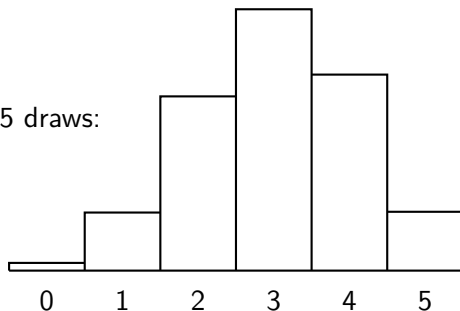
[Proof: A sequence SSFSFFS... has prob. $p^k(1 - p)^{n-k}$, by independence, if k is the # of S's. There are $\binom{n}{k}$ such sequences.]

Binomial Distribution

- ▶ **Eg:** Draw 5 balls from an urn, with 6 red balls and 4 green balls. $X = \#$ of reds in 5 draws.
If we draw *with replacement* then the draws are independent, and $X \sim \text{Bin}(5, 0.6)$. So $P(X = 2) = \binom{5}{2} 0.6^2 0.4^3$
- ▶ **Eg:** An opinion poll with yes/no answers will have a binomially distributed number of “yes” responses. (Provided it is done well, to ensure independence of responses).
- ▶ **Eg:** The number of girls among a family of 4 children is $\text{Bin}(4, \frac{1}{2})$ (ignoring the possibility of identical twins). So the probability of getting 2 boys and 2 girls is $\binom{4}{2} (\frac{1}{2})^2 (\frac{1}{2})^2 = \frac{3}{8} < \frac{1}{2}$. We'll see that a “balanced” family is the most likely single configuration, but families are more likely to be unbalanced.
- ▶ The binomial distribution is *unimodal*, ie probabilities go up and then down. Eg, histogram for the above urn example:

Mode of the binomial

Histogram, reds in 5 draws:

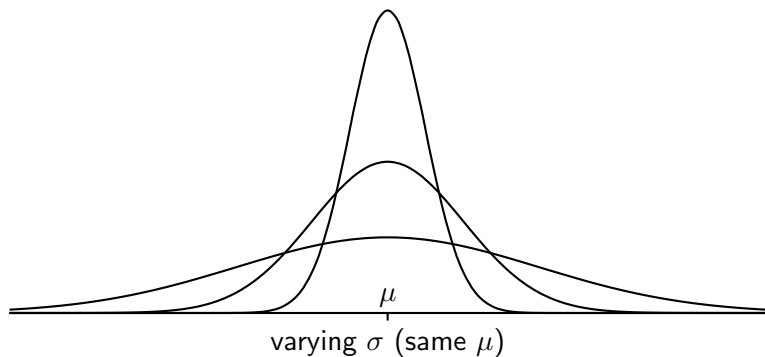
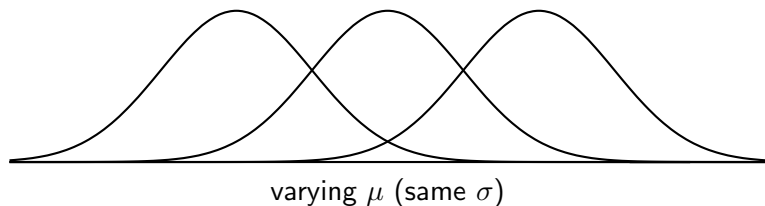


- ▶ The *Mode* of a distribution is the most likely value (there may be more than one mode, in case of ties). For $\text{Bin}(n, p)$ this is always either the integer immediately $\leq np$ or $\geq np$. The formula is that a mode is $= \lfloor (n+1)p \rfloor$, ie the greatest integer $\leq (n+1)p$. (And ties are possible).
- ▶ In the family example above, 2 is the mode. For families of 5 children, 2 and 3 are both modes.

Normal Distribution

- ▶ X has a *Normal* or *Gaussian* distribution, with parameters μ and σ^2 if its density is $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Here $\sigma > 0$ and μ is arbitrary. We write $X \sim N(\mu, \sigma^2)$.
- ▶ We will soon identify μ as the *mean* of the distribution, σ^2 as the *variance*, and σ as the *standard deviation*.
- ▶ But even now we see that μ is a location parameter (changing μ just shifts the distribution without changing its shape), and σ is a scale parameter (the distribution is concentrated around μ when σ is small, and is spread out when σ is big.)
- ▶ $N(\mu, \sigma^2)$ is unimodal, with mode at μ .

Normal Densities



Normal Density

We should check that the normal density really is a density, ie that it integrates to 1. The derivation uses material from MATH 2310, which is not part of this course (and you are not responsible for it). But I include it for completeness.

Change variables to $z = \frac{x-\mu}{\sigma}$. We want to show that $I = 1$, where

$$I = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Square this, convert it to a double integral, and then change variables to polar coordinates. We get

$$\begin{aligned} I^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{z^2+w^2}{2}} dz dw = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta \\ &= \frac{1}{2\pi} \begin{bmatrix} 2\pi & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \infty & -e^{-\frac{r^2}{2}} \\ 0 & 1 \end{bmatrix} = \frac{1}{2\pi} \times 2\pi \times 1 = 1. \end{aligned}$$

So $I = 1$, which is what we wanted to show.

Normal cdf

- ▶ Let $\Phi(z)$ be the cdf of a standard normal r.v. Z (ie $Z \sim N(0, 1)$ has $\mu = 0$ and $\sigma = 1$). $\Phi'(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$.
- ▶ There is no closed-form expression for Φ , so you have a table of values instead (Appendix 5).
- ▶ We calculate probabilities for Normal r.v.'s using Φ plus:
 - ▶ the general cdf formulae obtained earlier;
 - ▶ continuity (ie $\Phi(z-) = \Phi(z)$);
 - ▶ symmetry ($\Phi(-z) = P(Z \leq -z) = P(Z \geq z) = 1 - \Phi(z)$);
 - ▶ transformations (see below)
- ▶ **Lemma** If $X = \mu + \sigma Z$ then $X \sim N(\mu, \sigma^2) \Leftrightarrow Z \sim N(0, 1)$.
[Proof. Let $Z \sim N(0, 1)$ and $X = \mu + \sigma Z$. The cdf of X is $F(x) = P(X \leq x) = P(Z \leq \frac{x - \mu}{\sigma}) = \Phi\left(\frac{x - \mu}{\sigma}\right)$. So X has density $F'(x) = \frac{1}{\sigma}\Phi'\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x - \mu)^2}{2\sigma^2}}$.
The converse is similar.]

Normal probabilities

Eg: Let $X \sim N(1, 4)$. Find $P(0.5 \leq X \leq 3.46)$.

$\mu = 1$ and $\sigma^2 = 4$, so $\sigma = 2$. Therefore

$$\begin{aligned}P(0.5 \leq X \leq 3.46) &= P\left(\frac{0.5 - 1}{2} \leq \frac{X - 1}{2} \leq \frac{3.46 - 1}{2}\right) \\&= P(-0.25 \leq Z \leq 1.23) = \Phi(1.23) - \Phi(-0.25) \\&= \Phi(1.23) - (1 - \Phi(0.25)) = 0.8907 - 1 + 0.5987 \\&= 0.4894\end{aligned}$$

Here we've used the transformation $z = \frac{x-1}{2}$, continuity of Φ [so we didn't need $\Phi(-0.25-)$], symmetry, and have looked up two values from Appendix 5.

Normal probabilities

- ▶ What if the value $\Phi(z)$ you want isn't in the table? Basically your choices are
 - ▶ software
 - ▶ crude approximation (ie round z to 2 decimals and use the corresponding value from the table)
 - ▶ linear interpolation.
 - ▶ [And if programming, there are other useful approximation formulae, eg. on p. 95 of the text]
- ▶ The best answer (if you have a computer) is to use software. Eg, the NORMSDIST function in Excel computes the $N(0, 1)$ cdf for you. There are similar functions in all statistical software (eg R is a nice statistical package, that is free to download. In R the command is `pnorm`)
- ▶ Linear interpolation says that if $l \leq x \leq r$ and $x = l + \lambda(r - l)$ then $\Phi(x) \approx \Phi(l) + \lambda(\Phi(r) - \Phi(l))$. That is, $\Phi(x) \approx \Phi(l) + \frac{x-l}{r-l}(\Phi(r) - \Phi(l))$. This is exact for $x = r$ or $x = l$.

Normal probabilities

Eg: $X \sim N(2, 3)$. Find $P(X \leq 4)$.

$\mu = 2$ and $\sigma = \sqrt{3}$. So

$$P(X \leq 4) = P\left(\frac{X-2}{\sqrt{3}} \leq \frac{4-2}{\sqrt{3}}\right) = P(Z \leq 1.1547) = \Phi(1.1547).$$

- ▶ Most accurate: = NORMSDIST(1.1547)=0.87589
- ▶ Least accurate: $1.1547 \approx 1.15$ so
 $\Phi(1.1547) \approx \Phi(1.15) = 0.8749$
[Note a bad answer, but only accurate to 3 decimals]
- ▶ Reasonably accurate: $1.1547 = 1.15 + 0.47 \times (1.16 - 1.15)$ so
 $\Phi(1.1547) \approx \Phi(1.15) + 0.47 \times (\Phi(1.16) - \Phi(1.15))$
 $= 0.8749 + 0.47 \times (0.8770 - 0.8749) = 0.8758$
(now accurate to 4 decimals).

Normal probabilities

Eg: There is a rule of thumb that for normal distributions

- ▶ 70% of the mass lies within 1 standard deviation of the mean
ie. $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.70$
- ▶ 95% of the mass lies within 2 standard deviations of the mean
ie. $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$
- ▶ 99% of the mass lies within 3 standard deviations of the mean
ie. $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.99$

These round figures are easy to remember, but we can now calculate more refined answers, taking $\Phi(1)$, $\Phi(2)$, $\Phi(3)$ from the table. We would get:

- ▶ $P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) \approx 0.6827$
- ▶ $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) \approx 0.9545$
- ▶ $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) \approx 0.9973$

Normal approximation

- ▶ $\text{Bin}(n, p)$ prob's can be worked out exactly, when n is small.
- ▶ But when n is large, it is impractical to use the exact formulae. Instead, we approximate binomial probabilities by normal probabilities.
- ▶ For now, take the following as an empirical observation:
Let n be large and $X \sim \text{Bin}(n, p)$.
Then $X \approx Y$, where $Y \sim N(np, np(1 - p))$.
[We will see a rationale later, including the reason why we take $\mu = np$ and $\sigma^2 = np(1 - p)$.]
- ▶ This gives the following (crude) approximation formula:
 $X \sim \text{Bin}(n, p)$ and n large $\Rightarrow P(X \leq x) \approx P(Y \leq x)$
where $Y \sim N(np, np(1 - p))$.
- ▶ **Eg:** $X \sim \text{Bin}(1000, 0.5)$. Find $P(X \leq 495)$.
 $\mu = 1000 \times 0.5 = 500$, $\sigma^2 = 1000 \times \frac{1}{2} \times \frac{1}{2} = 250$.
So $P(X \leq 495) \approx P(Y \leq 495) = P(Z \leq \frac{495 - 500}{\sqrt{250}})$
 $= \Phi(-0.3162) = 0.3759$

Continuity Correction

- ▶ Note that this crude approximation gives $P(X = 495) \approx P(Y = 495) = 0$ since Y has a continuous distribution. It may be true that $P(X = 495)$ is small. But how small? Somehow we need to correct for approximating a discrete distribution by a continuous one.
- ▶ For a general discrete r.v. X , taking possible values x_1, \dots, x_n let $\delta_i = x_{i+1} - x_i$ be the distance between neighbouring values. Splitting the difference between neighbouring values, we have that x_i is the only possible value for X in the interval $[x_i - \frac{\delta_{i-1}}{2}, x_i + \frac{\delta_i}{2}]$ (Note: take $\delta_0 = -\infty$, $\delta_n = +\infty$). So $P(X \leq x_i) = P(X \leq x_i + \frac{\delta_i}{2})$, $P(X \geq x_i) = P(X \geq x_i - \frac{\delta_{i-1}}{2})$ and $P(X = x_i) = P(x_i - \frac{\delta_{i-1}}{2} \leq X \leq x_i + \frac{\delta_i}{2})$. If we're approximating X by a r.v. with a continuous distribution, we'll generally get more accurate answers if we apply the approximation to these expanded events (which are less sensitive to changing x) rather than the original ones.

Continuity Correction

- ▶ In the binomial case all the $\delta_i = 1$.
- ▶ **Eg:** $X \sim \text{Bin}(1000, 0.5)$. Then
$$\begin{aligned}P(X = 495) &= P(494.5 \leq X \leq 495.5) \\&\approx P(494.5 \leq Y \leq 495.5) = P\left(\frac{494.5-500}{\sqrt{250}} \leq Z \leq \frac{495.5-500}{\sqrt{250}}\right) \\&= \Phi(-0.2846) - \Phi(-0.3479) = 0.0240\end{aligned}$$
- ▶ **Eg:** $P(X \leq 495) = P(X \leq 495.5) \approx P(Y \leq 495.5)$
$$= P\left(Z \leq \frac{495.5-500}{\sqrt{250}}\right) = \Phi(-0.2846) = 0.3880$$

This will typically be a more accurate approximation than the cruder version given earlier.
- ▶ To summarize, there are multiple choices to make. We can do normal approximation with or without a continuity correction (but including the correction gives greater accuracy when approximating binomials). And the normal probabilities can be found using software, crude rounding, or linear interpolation.

Eg: Batting averages

- ▶ (§2.2 Problem 11a) If a player's true batting average is .300, what is the probability of hitting .310 or better over the next 100 at bats?
- ▶ Let X be the number of hits in 100 at bats. Assuming that at bats are independent, and that the probability of a hit is 0.3 for each at bat, we have $X \sim \text{Bin}(100, 0.3)$. $100 \times .310 = 31$, so we're asked for $P(X \geq 31)$.
- ▶ We don't want to work out the exact formula
$$\binom{100}{31}(.3)^{31}(.7)^{69} + \binom{100}{32}(.3)^{32}(.7)^{68} + \dots + \binom{100}{100}(.3)^{100}(.7)^0$$
- ▶ So approximate: $X \approx Y$ where $Y \sim N(\mu, \sigma^2)$ with $\mu = np = 30$ and $\sigma^2 = np(1-p) = 21$.
- ▶ The crudest answer would be
$$P(X \geq 31) \approx P(Y \geq 31) = P\left(\frac{Y-30}{\sqrt{21}} \geq \frac{31-30}{\sqrt{21}}\right)$$
$$= P(Z \geq .2182) = 1 - \Phi(.2182)$$
$$\approx 1 - \Phi(.22) = 1 - .5871 = .4129$$

Eg: Batting averages

- ▶ Interpolation is better:

$$\Phi(.2182) \approx \Phi(.21) + .82[\Phi(.22) - \Phi(.21)] = .5864 \text{ so}$$
$$P(X \geq 31) \approx 1 - .5864 = .4136$$

- ▶ And Excel is even better: $\text{NORMSDIST}(.2182) = .58637$ so
- $$P(X \geq 31) \approx 1 - .5864 = .41363$$

- ▶ But better than either of those improvements is incorporating the continuity correction.

$$P(X \geq 31) = P(X \geq 30.5) \approx P(Y \geq 30.5)$$
$$= P\left(\frac{Y-30}{\sqrt{21}} \geq \frac{30.5-30}{\sqrt{21}}\right) = P(Z \geq .1091) = 1 - \Phi(.1091)$$

Now crude rounding would give $\approx .4562$, and interpolation or Excel would both give $\approx .4566$

- ▶ In fact, using Excel one can compute the true value as being .4509 so in this case the continuity correction improves accuracy much more than interpolation, and brings the normal approximation to within 2% of the true answer.

Means and expected values

- ▶ There are multiple ways of identifying a “typical” or “average” value of a random variable X :
 - ▶ The *mode*: most likely value, ie the x or x 's maximizing $P(X = x)$ [discrete case] or the density $f(x)$ [continuous case];
 - ▶ The *median*: a value x (there may be more than one) such that $P(X \geq x) \geq \frac{1}{2}$ and $P(X \leq x) \geq \frac{1}{2}$. (In the continuous case, this simplifies to having the cdf $F(x) = \frac{1}{2}$.)
 - ▶ The *mean*: this is the right notion if we're dealing with *long-run averages*.
- ▶ **Def'n:** The *mean* or *expected value* of a r.v. X is $E[X] = \sum_{\text{values } x} xP(X = x)$ [discrete case], or $E[X] = \int_{-\infty}^{\infty} xf(x) dx$ [continuous case].
- ▶ **Note:** To be sure these sums & integrals make sense, we will always assume that X is *integrable*, ie that $\sum |x|P(X = x) < \infty$ or $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$.

Means

- ▶ In other words, $E[X]$ is a weighted average of the values, with the weights either probabilities or densities.
- ▶ Within a few weeks we will be able to prove the *Law of Large Numbers*, that says that if X_1, X_2, \dots are independent integrable r.v.'s, with the same distribution as X , then

$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow E[X]$$

in some sense, as $n \rightarrow \infty$.

- ▶ So, for example, if we repeatedly play some game, and X_i is how much we win or lose on the i th round, then over the long run, the amount we win or lose per round is the mean $E[X]$.

Means

- ▶ **Linearity:** Expectations are linear:

$$E[X + Y] = E[X] + E[Y] \text{ and } E[cX] = cE[X].$$

[pf: will do the latter, in the discrete case: x is a possible value for $X \Leftrightarrow cx$ is a possible value for cX . So

$$E[cX] = \sum_x cx \cdot P(cX = cx) = c \sum_x x \cdot P(X = x) = cE[X]]$$

- ▶ **Positivity:** $X \geq 0 \Rightarrow E[X] \geq 0$.

- ▶ **Eg:** $E[c] = c$

[pf: only one value, taken with probability 1, so

$$E[c] = c \times 1 = c.]$$

- ▶ **Eg:** Find $E[X]$ if

x	-1	0	1	3	5
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

$$E[X] = -1 \times \frac{1}{2} + 0 \times \frac{1}{4} + 1 \times \frac{1}{12} + 3 \times \frac{1}{12} + 5 \times \frac{1}{12} = \frac{3}{12} = \frac{1}{4}.$$

- ▶ **Eg:** If X is uniform on $\{x_1, x_2, \dots, x_n\}$, then $E[X] = \frac{x_1 + \dots + x_n}{n}$: the *arithmetic mean*.

Means

- ▶ **Eg:** $X \sim \text{Uniform on } [a, b]$.

$$\begin{aligned} \text{Then } E[X] &= \int_{-\infty}^{\infty} xf(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\ &= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}, \text{ the midpoint of the interval.} \end{aligned}$$

- ▶ **Eg:** $X \sim N(\mu, \sigma^2)$.

If $Z \sim N(0, 1)$ then $E[Z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-z^2/2} dz = 0$ by symmetry (the integrand is an odd function). So by linearity, $E[X] = E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu$.

- ▶ **Eg:** $X \sim \text{Bin}(n, p)$.

$$\begin{aligned} \text{1st approach: definition. } E[X] &= \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{1+k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!((n-1)-j)!} p^j (1-p)^{(n-1)-j} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} = np(p + (1-p))^{n-1} = np. \end{aligned}$$

Method of Indicators

- ▶ For an event A , define an *indicator random variable*

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A. \end{cases}$$

So $1 \leftrightarrow A$ occurs, $0 \leftrightarrow A$ doesn't occur.

- ▶ $E[1_A] = 0 \times P(A^c) + 1 \times P(A) = P(A)$.
- ▶ If A_1, \dots, A_n are events, and X counts the number which occur, then $X = \sum 1_{A_k}$ (adding up 0's and 1's \Leftrightarrow counting the 1's). So $E[X] = E[\sum 1_{A_k}] = \sum E[1_{A_k}] = \sum P(A_k)$.
- ▶ **Eg:** $X \sim \text{Bin}(n, p)$.

2nd approach: indicators. Let A_k be the event that the k th trial is a success. Then

$$E[X] = E[\sum_{k=1}^n 1_{A_k}] = \sum_{k=1}^n P(A_k) = \sum_{k=1}^n p = np.$$

Hypergeometric Mean

Eg: An urn has R red balls and Y yellow balls. Draw n without replacement, and let X count the number of reds [so X has a *hypergeometric* distribution].

- ▶ Let $N = R + Y$. We could work this out directly:

$E[X] = \sum_{k=0}^n k \times \frac{\binom{R}{k} \binom{Y}{n-k}}{\binom{N}{n}}$ if n is small. There's a similar expression for general $k \leq n$ except that one needs $0 \leq k \leq R$ and $0 \leq n - k \leq Y$ (otherwise we run out of balls).
Now cancel and simplify as in the binomial case ...

- ▶ Indicators are much easier: Let A_i be the event that the i th draw gives a red ball. By symmetry, $P(A_i) = \frac{R}{N}$ for each i .
So $E[X] = E[\sum_{i=1}^n 1_{A_i}] = \sum_{i=1}^n E[1_{A_i}] = \sum_{i=1}^n P(A_i) = \frac{nR}{N}$.

Variances

- ▶ The *variance* of X is $\text{Var}[X] = E[(X - E[X])^2]$.
If we approx. X by its mean, this = the mean-squared error.
- ▶ The *standard deviation* of X is $\text{SD}[X] = \sqrt{\text{Var}[X]}$.
The square root puts $\text{SD}[X]$ in the same units as X .
- ▶ Both measure the degree of uncertainty or randomness in X :
 $\text{Var}[X] = 0$ means X is constant.
- ▶ **2nd moment formula:** $\text{Var}[X] = E[X^2] - E[X]^2$.
Proof: $\text{Var}[X] = E[(X - E[X])^2]$
 $= E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2$
 $= E[X^2] - E[X]^2$.

Variances

► Other Properties:

1. $\text{Var}[aX + b] = a^2\text{Var}[X]$

Proof: $E[(aX + b - E[aX + b])^2] = E[(aX - aE[X])^2]$
 $= E[a^2(X - E[X])^2] = a^2\text{Var}[X].$

2. $\text{SD}[aX + b] = |a|\text{SD}[X].$

3. X_1, \dots, X_n independent $\Rightarrow \text{Var}[\sum X_k] = \sum \text{Var}[X_k].$

[We'll come back and prove in a week or so, after studying more about independence]

- To calculate $\text{Var}[X]$ we need to work out $E[X^2]$. We could do this by doing a transformation and finding the cdf of X^2 . But a simpler formula is available:

$$E[g(X)] = \sum_x g(x)P(X = x) \text{ (discrete case)}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx \text{ (continuous case)}$$

Proof: In the discrete case, let x_i be the values of X , and let A_i be the event that $X = x_i$. Then $g(X) = \sum_i g(x_i)1_{A_i}$, which gives the formula immediately.

Variances

In the continuous case, we'll only give the proof when g is smooth, increasing, 1-1, and onto. If $Y = g(X)$ and $h(y)$ is the density of Y , then $h(y) = f(x)/g'(x)$ [from transformations]. So

$E[Y] = \int_{-\infty}^{\infty} yh(y) dy = \int_{-\infty}^{\infty} g(x) \times \frac{f(x)}{g'(x)} \times g'(x) dx$, which gives the formula.

► Eg:

x	-1	0	1	3	5
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

We know from before that the mean $= \frac{1}{4}$. We could use

$$\text{Var}[X] = (-1 - \frac{1}{4})^2 \times \frac{1}{2} + (0 - \frac{1}{4})^2 \times \frac{1}{4} + (1 - \frac{1}{4})^2 \times \frac{1}{12} + (3 - \frac{1}{4})^2 \times \frac{1}{12} + (5 - \frac{1}{4})^2 \times \frac{1}{12}.$$

But the 2nd moment formula is better:

$$E[X^2] = \frac{(-1)^2}{2} + \frac{(0)^2}{4} + \frac{(1)^2}{12} + \frac{(3)^2}{12} + \frac{(5)^2}{12} = \frac{41}{12}.$$

$$\text{So } \text{Var}[X] = E[X^2] - E[X]^2 = \frac{41}{12} - \left(\frac{1}{4}\right)^2 = \frac{161}{48}$$

- **Eg:** $X \sim \text{Uniform on } [a, b]$:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{3(b-a)} \left[x^3 \right]_a^b \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}. \text{ So } \text{Var}[X] = E[X^2] - E[X]^2 \\ &= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

Of course, the smaller the interval, the smaller the variance.

- **Eg:** Normal $X \sim N(\mu, \sigma^2)$

Take $Z \sim N(0, 1)$ and integrate by parts.

$$\begin{aligned} E[Z^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} -ze^{-z^2/2} \right] + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 0 + 1 = 1. \end{aligned}$$

So by scaling, $\text{Var}[X] = \text{Var}[\mu + \sigma Z] = \sigma^2 \text{Var}[Z] = \sigma^2$.

In other words, we've basically used the mean and variance to parametrize $N(\mu, \sigma^2)$.

Binomial Variance

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

- ▶ We can find the variance directly:

$$\begin{aligned} E[X^2] &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n [k(k-1) + k] \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=2}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + E[X] \\ &= \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} + np \\ &= \sum_{k=2}^n \frac{n(n-1)(n-2)!}{(k-2)!((n-2)-(k-2))!} p^{2+k-2} (1-p)^{(n-2)-(k-2)} + np \\ &= n(n-1)p^2 \sum_{j=0}^{n-2} \frac{(n-2)!}{j!((n-2)-j)!} p^j (1-p)^{(n-2)-j} + np \\ &= n(n-1)p^2 + np \text{ by the binomial theorem.} \\ \text{So } \text{Var}[X] &= E[X^2] - E[X]^2 = n(n-1)p^2 + np - (np)^2 \\ &= np[(n-1)p + 1 - np] = np(1-p) \end{aligned}$$

Binomial/Hypergeometric Variance

- ▶ Or use indicators: $\text{Var}[1_A] = E[1_A^2] - E[1_A]^2 = E[1_A] - P(A)^2 = P(A) - P(A)^2 = P(A)[1 - P(A)]$.

So let A_i be the event that the i th trial is a success. If we jump ahead and use property 3 from (not proved yet), then by independence,

$$\text{Var}[X] = \text{Var}[\sum 1_{A_i}] = \sum \text{Var}[1_{A_i}] = \sum p(1-p) = np(1-p).$$

Eg: Hypergeometric variance.

- ▶ For notation, refer to the mean calculation.
- ▶ $X = \sum_{i=1}^n 1_{A_i}$, so $E[X^2] = E[\sum_{i,j=1}^n 1_{A_i} 1_{A_j}] = \sum_{i,j=1}^n E[1_{A_i} 1_{A_j}] = \sum_{i,j=1}^n P(A_i \cap A_j)$.
- ▶ If $i = j$ then $P(A_i \cap A_j) = P(A_i) = \frac{R}{N}$ by symmetry.
- ▶ If $i \neq j$ then $P(A_i \cap A_j) = \frac{R}{N} \times \frac{R-1}{N-1}$.
- ▶ So $E[X^2] = n \times \frac{R}{N} + n(n-1) \times \frac{R(R-1)}{N(N-1)}$.

Hypergeometric Variance

- ▶ Therefore $\text{Var}[X] = E[X^2] - E[X]^2$
$$= \frac{nR}{N} + \frac{n(n-1)R(R-1)}{N(N-1)} - \left(\frac{nR}{N}\right)^2 = \frac{nR}{N} \left(1 + \frac{(n-1)(R-1)}{N-1} - \frac{nR}{N}\right)$$
$$= \frac{nR}{N} \times \frac{N^2 - N + NnR - NR - Nn + N - nRN + nR}{N(N-1)} = \frac{nR}{N} \times \frac{(N-R)(N-n)}{N(N-1)}.$$
- ▶ We can interpret this by setting $p = \frac{R}{N}$, the probability of getting red on a single draw. Then $E[X] = np$ and $\text{Var}[X] = np(1-p)\frac{N-n}{N-1}$.
- ▶ In other words, the mean of X is the same, whether we sample with replacement (binomial) or without replacement (hypergeometric). But the variance gets SMALLER when we sample without replacement. The additional factor $\frac{N-n}{N-1}$ is known as a *finite size correction factor*.