Lecture 16: Mean and variance of random variables: Kelly 4.1-4.3

16.1: Expected value of a random variable: the mean

(i) Discrete case

If X is discrete with p.m.f. $P(X = x) = p(x) > 0$ for $x \in \mathcal{X}$, the *expected value* of X denoted $\mathrm{E}(X)$ is $\mathrm{E}(X) = \sum_{x \in \mathcal{X}} x\, p(x)$, provided this sum exists and is finite.

(ii) Continuous case

If X is continuous with p.d.f. $f(x)$, $f(x) \geq 0$ for $-\infty < x < \infty$. the *expected value* of X denoted $\mathrm{E}(X)$ is $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$, provided this integral exists and is finite. (Note $f(x)\, dx \approx P(x < X \leq x + dx)$.)

16.2: Expected value of a function of a random variable

(i) Discrete case

Now if X takes values x_i, the values taken by $Y = g(X)$ are $g(x_i)$, but several x_i may have the same $g(x_i)$.

$$P(Y = y) = \sum_{i: g(x_i) = y} p(x_i) \quad \text{so}$$

$$\mathrm{E}(g(X)) = \sum_y y\, P(Y = y) = \sum_y \left(y \sum_{i: g(x_i) = y} p(x_i) \right) = \sum_y \sum_{i: g(x_i) = y} g(x_i) p(x_i) = \sum_i g(x_i) p(x_i).$$

(ii) Continuous case (not proved)

For a continuous random variable $\mathrm{E}(g(X) = \int_x g(x)\, f(x)\, dx$.

(iii) An important property (proved for discrete random variables):

$$\mathrm{E}(g_1(X) + g_2(X)) = \sum_x (g_1(x) + g_2(x)) p(x) = \sum_x g_1(x)\, p(x) + \sum_x g_2(x) p(x) = \mathrm{E}(g_1(X)) + \mathrm{E}(g_2(X))$$

(iv) A simple property

$$\mathrm{E}(aX + b) = \sum_x (ax + b) p(x) = a \sum_x x\, p(x) + b \sum_x p(x) = a\, \mathrm{E}(X) + b$$

(v) **Note:** The same results (ii) and (iii) holds for continuous random variables.

16.3: The variance of a random variable

(i) Definition: If $E(X) = \mu$, $\mathrm{var}(X) = \mathrm{E}(X - \mu)^2$

Since $(x - \mu)^2 \geq 0$ for every x, the definition shows $\mathrm{var}(X) \geq 0$.

(ii) Property 1: using 16.2 (ii) we have

$\mathrm{var}(X) = \mathrm{E}(X - \mu)^2 = \mathrm{E}(X^2 - 2\mu X + \mu^2) = \mathrm{E}(X^2) - 2\mu \mathrm{E}(X) + \mu^2 = \mathrm{E}(X^2) - 2\mu \times \mu + \mu^2$
$= \mathrm{E}(X^2) - \mu^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$. (This is usually the easiest way to compute $\mathrm{var}(X)$.)

(iii) Property 2: using 16.2 (ii) we have

$$\mathrm{var}(aX + b) = \mathrm{E}((aX + b - a\mu - b)^2) = \mathrm{E}(a^2 (X - \mu)^2) = a^2 \mathrm{E}((X - \mu)^2) = a^2 \mathrm{var}(X).$$

16.4 One example: the mean and variance of a Poisson random variable $P(X = x) = e^{-\lambda} \lambda^x / x!$.

The mean: $\mathrm{E}(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \lambda^x / x! = \sum_{x=1}^{\infty} x e^{-\lambda} \lambda^x / x! = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \lambda^{x-1} / (x-1)! = e^{-\lambda} \lambda e^{\lambda} = \lambda$

Then: $\mathrm{E}(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) e^{-\lambda} \lambda^x / x! = e^{-\lambda} \lambda^2 \sum_{x=2}^{\infty} \lambda^{x-2} / (x-2)! = e^{-\lambda} \lambda^2 e^{\lambda} = \lambda^2$

So then $\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - (\mathrm{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Lecture 17: More expectations. Summing independent random variables

(Here we take the simplest case of two discrete random variables, but the results are true in general.)

17.1 Independent random variables

Let X and Y be two discrete random variables defined on the same sample space. (e.g. Ω is outcomes of tosses of two dice; X is the sum of the two values, Y is the maximum of the two values.)

Suppose X takes values x_i, $i = 1, 2, ...,$ and Y takes values y_j, j=1,2,3,....

Let $P(X = x_i \cap Y = y_j) = p_{ij}$. Note $\sum_i \sum_j p_{ij} = 1$.

Also $P(X = x_i) = \sum_j P(X = x_i \cap Y = y_j) = \sum_j p_{ij}$

and $P(Y = y_j) = \sum_i P(X = x_i \cap Y = y_j) = \sum_i p_{ij}$.

Also note that *events* of interest are of the form $(X = x_i \cap Y = y_j)$; all other statements about X and Y derive from these.

Definition: Random variables X and Y are **independent** if

$$p_{ij} = P(X = x_i \cap Y = y_j) = P(X = x_i) \times P(Y = y_j) \text{ for all } i \text{ and } j.$$

17.2 Expectation of the sum (Note: this does not require independence.)

The set of possible values of $X + Y$ is the set of all $x_i + y_j$ for all i and j:

$$\begin{aligned} \mathrm{E}(X+Y) &= \sum_i \sum_j (x_i + y_j) P(X = x_i \cap Y = y_j) = \sum_i \sum_j x_i p_{ij} + \sum_i \sum_j y_j p_{ij} \\ &= \sum_i (x_i \sum_j p_{ij}) + \sum_j (y_j \sum_i p_{ij}) = \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j) \\ &= \mathrm{E}(X) + \mathrm{E}(Y) \quad \textbf{Expectation of sum is sum of expectations: always}. \end{aligned}$$

17.3 Expectation of the product: independence case

In general: $\mathrm{E}(XY) = \sum_i \sum_j x_i y_j P(X = x_i \cap Y = y_j) = \sum_i \sum_j x_i y_j p_{ij}$.

If X and Y are **independent** then

$$\begin{aligned} \mathrm{E}(XY) &= \sum_i \sum_j x_i y_j p_{ij} = \sum_i \sum_j x_i y_j P(X = x_i) P(Y = y_j) \\ &= \left(\sum_i x_i P(X = x_i) \right) \left(\sum_j y_j P(Y = y_j) \right) = \mathrm{E}(X)\mathrm{E}(Y). \end{aligned}$$

17.4 Variance of the sum: independence case

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

If X and Y are **independent**, $\mathrm{E}(XY) = \mathrm{E}(X).\mathrm{E}(Y)$. 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 1: If we can sum 2, we can sum any finite number: $X + Y + Z = (X + Y) + Z$.

Note 2: The converse of 17.3, 17.4 is NOT true.

We can have $\mathrm{E}(XY) = \mathrm{E}(X).\mathrm{E}(Y)$, but X and Y NOT **independent**.

Lecture 18: The Bernoulli process, and associated random variables (Kelly 1.4)

18.1: The process

0 0 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1: Each *trial* is *success* (1) or not (0).

$X_1 X_2 X_3 X_4$.... ...X_{25}...: Each X_i is 0 or 1.

... ... T_5... ... ... T_{10}... ... ... T_{15}... ... ... T_{20}... ... ... T_{25}...: $T_n = X_1 + ... + X_n$

$---Y_1 ----Y_2 Y_3 --Y_4 ----------Y_5 --Y_6 ---Y_7$: Y_r is rth inter-arrival time

... ...W_1... ... ...$W_2 W_3$.. ..W_4... ... ...W_5... ...W_6... ...W_7: W_r is total waiting time to rth 1.

The Bernoulli process is defined by X_i independent, with $P(X_i = 1) = p$ and $P(X_i = 0) = (1 - p)$.

$T_n = X_1 + ... + X_n$ is number of *successes* (i.e. 1s) in first n trials.

Y_r is the *inter-arrival time*: number of trials from $(r-1)$th success to r th.

$W_r = Y_1 + Y_2 + ... + Y_r$ is number of trials to r th. success.

$Y_r^* = Y_r - 1$: number of failures (0) before next success (1).

$W_r^* = Y_1^* + ... + Y_r^*$: number of failures (0) before r th success. **Note:** $W_r > n$ if and only if $T_n < r$.

18.2: Bernoulli and Binomial random variables

(i) X_i is Bernoulli(p). $P(X_i = 1) = p$, $P(X_i = 0) = (1 - p)$.

$\mathrm{E}(X_i) = p \times 1 + (1-p) \times 0 = p$

$\mathrm{E}(X_i^2) = p \times 1^2 + (1-p) \times 0^2 = p$, so $\mathrm{var}(X_i) = \mathrm{E}(X_i^2) - (\mathrm{E}(X_i))^2 = p - p^2 = p(1-p)$.

$T_n = X_1 + ... + X_n$ is Binomial (n, p).

The probability of each sequence of k 1's and $(n-k)$ 0's is $p^k (1-p)^{n-k}$ and there are $\binom{n}{k}$ such sequences.

$P(T_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Expectations always add: see 16.2. $\mathrm{E}(T_n) = \mathrm{E}(X_1) + ... + \mathrm{E}(X_n) = p + p + ... + p = np$.

The variances also add **because the trials are independent**:

$\mathrm{var}(T_n) = \mathrm{var}(X_1) + ... + \mathrm{var}(X_n) = np(1-p)$.

18.3: Geometric and Negative Binomial random variables

Y_r are independent, and have Geometric (p) distribution: $P(Y = k) = (1-p)^{k-1} p$, for $k = 1, 2, 3, ...$

$\mathrm{E}(Y) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = p/(1-(1-p))^2 = 1/p$.

$\mathrm{E}(Y(Y-1)) = 2p(1-p)/(1-(1-p))^3 = 2(1-p)/p^2$.

So $\mathrm{var}(Y) = \mathrm{E}(Y(Y-1) + Y) - (\mathrm{E}(Y))^2 = (2(1-p)/p^2) + 1/p - 1/p^2 = (1-p)/p^2$ (see "note").

$Y^* = (Y-1)$, $P(Y^* = k) = (1-p)^k p$, for $k = 0, 1, 2, 3....$

$\mathrm{E}(Y_r^*) = \mathrm{E}(Y) - 1 = (1-p)/p$, $\mathrm{var}(Y^*) = \mathrm{var}(Y)$; see 16.2,16.3.

Note: $\sum_r x^r = 1/(1-x)$, so $\sum_r r x^{r-1} = \frac{d}{dx}\frac{1}{1-x} = 1/(1-x)^2$, $\sum_r r(r-1)x^{r-2} = \frac{d}{dx}\frac{1}{(1-x)^2} = 2/(1-x)^3$.

$W_r = Y_1 + ... + Y_r$. Expectations add, so $\mathrm{E}(W_r) = r/p$.

Again the variances add, **because the intervals Y_i are independent**: $\mathrm{var}(W_r) = r(1-p)/p^2$.

$P(W_r = k) = P(r-1$ successes in $k-1$ trials, and then success$) = \binom{k-1}{r-1}(1-p)^{k-r} p^{r-1} p$ for $k = r, r+1, ...$

$W_r^* = W_r - r$, $\mathrm{E}(W_r^*) = \mathrm{E}(W_r) - r$, $\mathrm{var}(W_r^*) = \mathrm{var}(W_r)$

$P(W_r^* = k) = P(r-1$ successes in $r+k-1$ trials, and then success$) = \binom{r+k-1}{r-1}(1-p)^k p^{r-1} p$

for $k = 0, 1, 2, 3...$

Lecture 19: Examples of Binomial, Geometric and Negative Binomials

A hypothetical story:

Mendel crossed two plants that were red-flowered, but each had one white-flowered parent. He therefore knew that each offspring plant would have white flowers with probability 1/4, independently of all the others. He planted one offspring seed each morning, and they all grew, and each one flowered the exact same number of days after planting. The first one flowered on June 1, 1865.

1. By June 20, 20 plants had flowered.

(a) How many of these plants are expected to have white flowers?

(b) What is the variance of the number of white-flowered plants?

(c) What is the probability that 5 of the plants had white flowers?

2. On June 8th, Mendel saw that the plant newly flowered that day had white flowers.

(a) What is the expected date of the next white-flowered plant?

(b) What is the probability that the next white-flowered plant flowers June 15 *or later*?

(c) What is the probability the next white-flowered plant flowers on June 13?

3. On June 8th, Mendel saw that the plant newly flowered that day had white flowers.

(a) What is the expected number of red flowers flowering before the next white-flowered plant?

(b) What is the variance of this number?

(c) What is the probability this number is at least 3?

(d) Mendel's assistant reminds him that the plant flowering on June 7 also had white flowers, and says that therefore by "the law of averages" they will probably have to wait longer than four days for the next white-flowered plant. What does Mendel say?

4. On June 8th, Mendel saw that the plant newly flowered that day had white flowers.

(a) What is the expected date of flowering of the 5 th white-flowered plant *after the one on June 8*?

(b) What is the probability the 5 th white-flowered plant *after the one on June 8* flowers on June 28?

Lecture 20: Poisson random variables: approximation to Binomial

20.1 Reminder of facts about the Poisson distribution

(i) From 11.3: $P(X = j) \ = \ e^{-\lambda}\lambda^j/j!, \quad$ for $j = 0, 1, 2, 3, ...$

(ii) From 16.4: $\mathrm{E}(X) \ = \ = \sum_{j=0}^{\infty} je^{-\lambda}\lambda^j/j! \ = \ \sum_{j=1}^{\infty} je^{-\lambda}\lambda^j/j! \ = \ e^{-\lambda}\lambda\sum_{j=1}^{\infty}\lambda^{j-1}/(j-1)! \ = \ e^{-\lambda}\lambda e^{\lambda} \ = \ \lambda$

Then: $\mathrm{E}(X(X-1)) \ = \ \sum_{j=0}^{\infty} j(j-1)e^{-\lambda}\lambda^j/j! \ = \ e^{-\lambda}\lambda^2\sum_{j=2}^{\infty}\lambda^{j-2}/(j-2)! \ = \ e^{-\lambda}\lambda^2 e^{\lambda} \ = \ \lambda^2$

So then $\mathrm{var}(X) \ = \ \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \ = \ \mathrm{E}(X(X-1)) \ + \ \mathrm{E}(X) - (\mathrm{E}(X))^2 \ = \ \lambda^2 + \lambda - \lambda^2 \ = \ \lambda$

(iii) Useful model for numbers of things (accidents, hurricanes, centenarians, errors, customers, judicial vacancies, ...), when there are a very large number of opportunities for the “thing” but each has small probability.

(iv) In these examples, the expected number of accidents, errors, judicial vacancies is “moderate” ($\mathrm{E}(X) \ = \ \lambda$). Typically λ is between 1 and 20. However, there is no hard upper bound.

20.2 Reminder of facts about the Binomial distribution

(i) From 11.3: $P(T = j) \ = \ \binom{n}{j} p^j(1-p)^{n-j}$, for $j = 0, 1, 2, ..., n$.

(ii) But also from 18.2, $T \ = \ X_1 + X_2 + ... + X_n$ where X_i are Bernoulli(p); $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$. Then $\mathrm{E}(X_i) \ = \ p$ and $\mathrm{var}(X_i) = p(1-p)$, so $\mathrm{E}(T) \ = \ np$ and $\mathrm{var}(T) \ = \ np(1-p)$.

Note, to sum the variances we are using the independence of the X_i.

(iii) Model for number of times something happens in n independent trials (coin tosses, red-flowered offspring pea-plants, Danes speaking German....) when the probability of the “thing” happening on each trial is p.

(iv) In these examples, p is “moderate” (0.25, 0.5, ...) and n also usually “moderate” (10 coin tosses, grow 30 pea plants, ...). There is a hard upper bound (n) on the value of X.

20.3 Poisson approximation to the Binomial

(i) Let X be a Binomial ($\mathrm{Bin}(n, p)$) random variable, and Y a Poisson random variable with parameter λ.

(ii) Suppose n gets large, and p gets small in such a way that np remains “ moderate”. Then we can match up the means: $\mathrm{E}(X) \ = \ np \ = \ \lambda \ = \ \mathrm{E}(Y)$.

(iii) Now $\mathrm{var}(X) \ = \ np(1-p) \ = \ \lambda(1-p) \ \approx \ \lambda \ = \ \mathrm{var}(Y)$.

(iv) In fact,

$$P(X = j) \ = \ \frac{n!}{j!(n-j)!}p^j(1-p)^{n-j} \ = \ \frac{n!}{j!(n-j)!}\left(\frac{\lambda}{n}\right)^j\left(1 - \frac{\lambda}{n}\right)^{n-j}$$

$$= \ \frac{n(n-1)...(n-j+1)}{n^j}\frac{1}{j!}\left(\frac{\lambda}{(1-\lambda/n)}\right)^j(1-\lambda/n)^n \ \approx \ 1\frac{1}{j!}\lambda^j\exp(-\lambda) \ = \ P(Y = j)$$

20.4 Back to the class data on birthdays Sample 1: $n = 31$, 2 pairs, 1 trio. Sample 2: $n = 31$, 2 pairs. Combined: $n = 62$, 6 pairs, 1 trio.

(i) Actual probability of no pairs in 31: $365 \times 364 \times ..... \times 335/(365)^{31} \ = \ 0.2695$. For 62 birthdays, probability of no pairs is 0.004. So it is not surprising we had pairs, but how many pairs should we get?

APPROXIMATION: $m = 31 \times 30/2 = 465$ not-quite-independent pairs, each pair probability $p = 1/365$.

$\lambda = 465/365 = 1.274$, $P(X = 0) \ = \ 0.2797$. $P(X = 1) \ = \ 1.274 \times 0.2797 = 0.3653$.

$P(X \geq 2) \ = \ 1 - P(X = 0) - P(X = 1) \ = \ 0.364$. So it was not surprising to get 2 pairs.

(ii) What about the trio in sample 1? Now $m \ = \ 31 \times 30 \times 29/6 \ = \ 4495$ trios, each with probability $p \ = \ 1/365^2 \ = \ 7.5 \times 10^{-6}$. So approximate by Poisson with mean $\lambda = 4495p \ = \ 0.0337$. Now $P(X = 0) \ = \ \exp(-0.0337) = 0.967$. So it was quite surprising to see a trio in a sample size 31 (about 3% chance).

(iii) What about a trio in the combined set of 62 birthdays? Now $m = 62 \times 61 \times 60/6 \ = \ 37820$ trios, so $\lambda \ = \ 37820/365^2 \ = \ 0.284$, and $\exp(-0.284) = 0.75$, so the chance of getting at least 1 trio on the set of 62 is about 25%.