17/6/21

Agenda: • Assignment 7 questions & answers
• Some general talk around mean & variance.

(Q1) • $\mathbb{E}[X] = \sum_{n=1}^{\infty} n\, \mathbb{P}(X=n)$.

• $\mathrm{Var}(X) = \mathbb{E}[|X-\mathbb{E}[X]|^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(a) $\Omega = \{1,2,\dots,n\}$, $p(x) = 1/n \ \forall x$,
$X(j) = j$, $j \in \Omega$
$\Rightarrow \mathbb{E}[X] = \sum_{j=1}^{n} j/n = \frac{n(n+1)}{2n} = \frac{n+1}{2}$.

$\mathbb{E}[X^2] = \sum_{j=1}^{n} j^2/n = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6}$.

$\Rightarrow \mathrm{Var}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = (n+1)\left[\frac{2n+1}{6} - \frac{n+1}{4}\right]$
$= (n+1)\left[\frac{n-1}{12}\right] = \left[\frac{n^2-1}{12}\right]$

(b) $\Omega = \{(a_1,a_2,\dots,a_n): a_i \in \{0,1\}\}$, $p(\omega) = 1/2^n \ \forall \omega \in \Omega$
$X(\omega) = \#\{i : a_i = 1\} \Rightarrow p(X=k) = \binom{n}{k} 2^{-n}$, $0 \le k \le n$
Let $X_i(\omega) = \mathbb{1}_{\{a_i=1\}} \Rightarrow X_i$ are independent and $X = \sum_{i=0}^{m} X_i$.
$\Rightarrow \mathrm{Var}(X) = \sum_{i=0}^{n} \mathrm{Var}(X_i)$.

$\mathbb{E}[X_i] = \mathbb{P}(X_i=1) = p$; $\mathbb{E}[X_i^2] = \mathbb{P}[X_i=1] = p$
$\Rightarrow \mathrm{Var}[X_i] = p(1-p)$
$\Rightarrow \mathrm{Var}(X) = np(1-p)$.

(c) Description of experiment:
- Collection of n objects
- Of which m are good.
- Draw r at a time, and count the no. of successes.

To write this clearly, let $1,\dots,m$ be "good" and $m+1,\dots,n$ be "bad". Then, we have the sample space
$\Omega = \{S \subseteq [n] : |S| = r\}$; $\mathbb{P}(S) = 1/\binom{n}{r}$.
$X(S) = \#\{i \in S : i \le m\}$.
Then, $\mathbb{P}(X=t) = \frac{\binom{m}{t}\binom{n-m}{r-t}}{\binom{n}{r}}$
$\forall \max\{0, r+m-N\} \le t \le \max\{m, r\}$

Idea: For $1 \le i \le m$, $X_i = \mathbb{1}_{\{i \in S\}}$. Then, $\sum_{i=1}^{m} X_i = X$, note that X_i need not be ind. b/c it is sampling without replacement.

However, $\mathbb{E}[X_i] = \mathbb{P}(i \in S) = \frac{r}{n}$; so $\mathbb{E}[X] = \frac{mr}{n}$.

$\mathbb{E}[X_i^2] = \mathbb{P}(i \in S) = \frac{mr}{n}$.

$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}[(\sum X_i)^2] - (\sum \mathbb{E}[X_i])^2$
$= \sum_{i,j} \mathbb{E}[X_i X_j] - \sum_{i,j} \mathbb{E}[X_i]\,\mathbb{E}[X_j]$.
$= \sum_{i=1}^{m} (\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2) + \left(\sum_{i \ne j} \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]\right)$
$= \frac{mr}{n}\left(1 - \frac{mr}{n}\right) m +$ "

Now, for $i \ne j$,
$\mathbb{E}[X_i X_j] = \mathbb{P}(i,j \in S) = \binom{n-2}{r-2} / \binom{n}{r} = \frac{r}{n} \times \frac{r-1}{n-1}$.

The number of $i \ne j$ is $\frac{m(m-1)}{2}$ so we get:

$= \frac{mr}{n}\left[1 - \frac{mr}{n}\right] m + \frac{m(m-1)}{2}\left[\frac{r(r-1)}{n(n-1)} - \frac{r^2}{n^2}\right]$.
Can be simplified if necessary.

(Q3) $\Omega = [6] \times [6]$, $p(\omega) = 1/36 \ \forall \omega \in \Omega$.
(a) $X((a,b)) = a+b$
Expand in detail to notice
$\mathbb{P}(X=2) = \mathbb{P}(X=12) = 1/36$, $\mathbb{P}(X=3) = \mathbb{P}(X=11) = 2/36$,
$\mathbb{P}(X=4) = \mathbb{P}(X=10) = 3/36$, $\mathbb{P}(X=5) = \mathbb{P}(X=9) = 4/36$,
$\mathbb{P}(X=6) = \mathbb{P}(X=8) = 5/36$, $\mathbb{P}(X=7) = 6/36$
i.e. $\mathbb{P}(X=a) = \frac{6 - |7-a|}{36}$, $a \in \{2,3,\dots,12\}$.

Note that $X = X_1 + X_2$, where $X_1(a,b) = a$
$X_2(a,b) = b$
& X_1, X_2 are ind. $\Rightarrow \mathbb{E}[X] = 3.5 + 3.5 = 7$.
Note $X_1, X_2 \sim \mathrm{Unif}[6]$
$\Rightarrow \mathrm{Var}(X) = 2\left(\frac{6^2-1}{12}\right) = 35/6$.

(b) $X((a,b)) = \max\{a,b\}$.
$\Rightarrow \mathbb{P}(X=k) = \frac{2k-1}{36}$, $k = 1,2,\dots,6$.
$\mathbb{E}[X] = \sum_{k=1}^{6} \frac{k(2k-1)}{36} = \frac{1}{36}\sum_{k=1}^{6}(2k^2 - k)$
$= \frac{1}{36}\left[\frac{6 \times 7 \times 13}{3} - 21\right] = \frac{1}{36}[161] = 161/36 \approx 4\frac{17}{36}$
$\mathbb{E}[X^2] = \sum_{k=1}^{6} \frac{k^2(2k-1)}{36} = \frac{1}{36}\left[\sum_{k=1}^{6}(2k^3 - k)\right]$
$= \frac{1}{36}\left[\frac{36 \times 49}{2} - 21\right] = \frac{1}{36}[861] = 861/36$.
$\Rightarrow \mathrm{Var}(X) = 861/36 - (161/36)^2 = 5075/1296$.

(Q5) $\Omega = \{(r_1,\dots,r_n): r_i \ge 0, \sum r_i = r\}$.
$\mathbb{P}((r_1,\dots,r_n)) = \frac{1}{n^r}\frac{r!}{r_1!\,r_2!\cdots r_n!}$
$\mapsto X((r_1\dots r_n)) = \#\{i : r_i = 0\}$.
Write $X_i = \mathbb{1}_{\{r_i=0\}}$, then $X = \sum_{i=1}^{n} X_i$.
Now, we will find
$\mathbb{P}(X_{i_1}=1,\dots,X_{i_j}=1) = \mathbb{P}(r_{i_1}=0,\dots,r_{i_j}=0)$
$= \sum_{\substack{r_i \ge 0, \sum r_i = n \\ r_{i_j} = 0 \forall j}} \frac{1}{n^r}\frac{r!}{r_1!\cdots r_n!} = \sum_{\substack{r_i \ge 0, \sum r_i = n \\ i = 0 \dots n-j}} \frac{1}{n^r}\frac{n!}{r_1!\cdots r_{n-j}!}$
$= \frac{1}{n^r}(n-j)^r$ by the multinomial thm.

Now use the generalized IEP:
$\{X=k\} = \{$For exactly k values of i, $X_i = 1\}$
$= \sum_{j=k}^{n} (-1)^{j-k}\binom{j}{k}\binom{n}{j}\frac{(n-j)^r}{n^r}$.

This gives us the pmf of X. However, we can get the expectation from:
$\mathbb{E}[X] = \sum \mathbb{E}[X_i] = \sum \mathbb{P}(X_i=1) = n\left(\frac{n-1}{n}\right)^r$

And the variance from:
$\mathbb{E}[X^2] = \sum \mathbb{E}[X_i^2] + \sum_{i \ne j} \mathbb{E}[X_i X_j]$
$= \sum \mathbb{P}[X_i=1] + \sum_{i \ne j} \mathbb{P}[X_i=1, X_j=1]$
$= n\left(\frac{n-1}{n}\right)^r + 2\binom{n}{2}\left(\frac{n-2}{n}\right)^r$

and finish.

(Q7) This is fairl com-ortable: Let X take distinct values $\{a_1, a_2, \dots, a_n\}$; $a_i \ge 0$. Let t be positive. Let $a_0 = 0$, $a_{n+1} = \infty$ (formally). Then, for t positive, we know $a_i \le t \le a_{i+1}$ for some $i \in \{0,\dots,n\}$. So, we get:
$t\,\mathbb{P}(X \ge t) \le \sum_{j=i+1}^{n} \mathbb{P}(X = a_{i+1})\,a_{i+1} \le \mathbb{E}[X]$
from where Markov follows. Chebyshev follows from:
$\mathbb{P}[|X - \mathbb{E}[X]| \ge t] = \mathbb{P}[(X - \mathbb{E}(X))^2 \ge t^2] \le \frac{\mathbb{E}[|X - \mathbb{E}(X)|^2]}{t^2}$
$= \mathrm{Var}(X)/t^2$

and done!

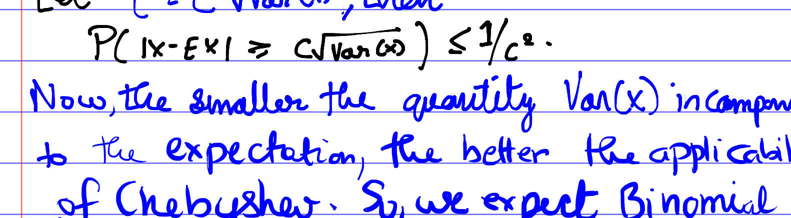
A look at variance - mean comparison.

The strength of Chebyshev's inequality is measured by the amount of variance of X. Ideally, X is expected to concentrate around its mean, if the variance is of order smaller than the mean.

Two examples
→ $U([n])$ — $\mathbb{E}[X] \approx n$, $\mathrm{Var}[X] \approx n^2$.

→ $\mathrm{Bin}(n,p)$ — $\mathbb{E}[X] \approx n$, $\mathrm{Var}(X) \approx n$.

If you look at the Chebyshev inequality,
$\mathbb{P}(|X - \mathbb{E}(X)| \ge t) \le \mathrm{Var}(X)/t^2$.
Let $t = c\sqrt{\mathrm{Var}(X)}$, then
$\mathbb{P}(|X - \mathbb{E}X| \ge c\sqrt{\mathrm{Var}(X)}) \le 1/c^2$.
Now, the smaller the quantity $\mathrm{Var}(X)$ in comparison to the expectation, the better the applicability of Chebyshev. So, we expect Binomial to concentrate better than a uniform.
Indeed, from the pmfs,

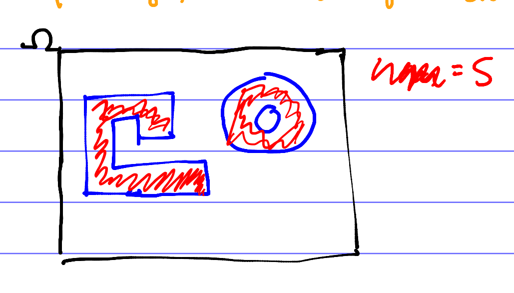


it is clear that the binomial is likely to be much closer to its mean than that of a uniform.

The point: knowing just the exp. & variance gives a great idea of the concentration of the r.v., without even knowing the pmf.

Briefly Monte-Carlo

Imagine you have a shape whose area you want to find. Let's say it's a subset of a unit square.



The idea is to have a shooter, whose job is to shoot bullets at the target, in a manner such that they are uniformly spread over Ω. Here's a question: If N is the number of shots, & n is the no. of shots that land inside S. Is $\frac{n}{N} \approx \mathrm{Area}(S)$?

Ans: Yes. That is the content of the laws of large numbers, and forms the basis of the Monte Carlo method.

Qn: How large an N do I need, to get a good idea of the Area of S?
Ans: Depends upon the Variance of the uniform random variable.
In general, a "Monte Carlo" scheme involves calculating a deterministic quantity, using independent identically distributed random variables. The convergence of the Monte-Carlo scheme involves the mean-variance relation of the random variable.
So, a MC scheme involving a Binomial r.v. is likely to converge faster & be accurate quicker than one involving a uniform r.v.

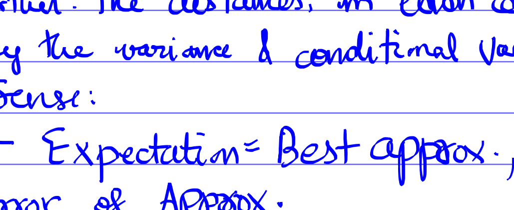
The idea of the mean and variance

Suppose you have points scattered on a plane, at positions v_i with mass m_i. Think of this as a "rigid" mass.
Qn: Imagine I had to interpret this as a "point" mass. Which point is the best for this purpose?

Answer: You try to minimize a "reasonable" distance function like so:
$f(x) = \sum m_i |x - v_i|^2$
This is the "squared distance" function.
Fact: If X is a random variable with $\mathbb{P}(X=v_i) \propto m_i$, then $\mathbb{E}[X]$ is the unique minimizer of f. In other words, $\mathbb{E}[X]$ is the best place for the whole rigid mass to be placed as a point mass. This is the "geometric" interpretation of $\mathbb{E}[X]$.
Translation: Without knowing anything about a random variable, the best guess you can take about its values, is its expectation.

If we know more, can we do better though? For functions $f, g: \{1,\dots,n\} \to \mathbb{R}$, define the "squared distance"
$\|f-g\| = \sum (f(x) - g(x))^2$
If we treat a function as a point, by taking $f(x) = (f(1),\dots,f(n)) \in \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{n \text{ times}}$,
then the distance between functions is given a geometric meaning as the distance between 2 points in a high dimensional space. So pmfs, are also "points" in a high dimensional space. With this, "geometrically" speaking:

If X is a random variable with pmf f, then conditioning X w.r.t. some other event, introduces a "set" S of possible pmfs that can be the conditional pmf. Now, the actual cond. pmf will be given by:
$f_c(x) = \underset{g \in S}{\mathrm{argmin}}\ |f - g|$
That is geometrically, the closest point in S, to f, is the conditional pmf!



Further: The distances, in each case, are given by the variance & conditional variance. So, in a sense:
— Expectation = Best approx., Variance = Error of Approx.
— Condnl Expn = Best approx. when the pmfs are restricted to a certain set, & Condnl Variance = Error of this Approx.