On Discrete Distributions

Tristan Chaang

February 13, 2022

In this article I will write about discrete distributions, including their properties and their derivations.

Contents

1 PGF

If the random variable is said to have a discrete probability distribution, it can only take countably many values. We simplify our analysis to studying $\mathbb{N}_0$. If X has a discrete distribution, it has a probability of $p_n = \mathcal{P}(X = n)$ of taking the value n $(n = 0, 1, \cdots)$. We can compile all of these values into a list:

n	0	1	2	3	4	5	6	7	8	$\cdots$
$\mathcal{P}(X = n)$	0.3	0.1	0.4	0.1	0.05	0.05	0	0	0	$\cdots$

Note that the sum of probabilities in the table must sum to 1. Another way of writing the list is by using *probability generating functions (pgf)*. This is a formal series taking $P(X = n)$ as coefficient of x^n. For the distribution above, the pgf is

$$G_X(x) = 0.3 + 0.1x + 0.4x^2 + 0.1x^3 + 0.05x^4 + 0.05x^5.$$

Definition. The probability generating function of X is

$$G_X(x) = \sum_{\forall n} p_n x^n$$

where $p_n = \mathcal{P}(X = n)$.

One might ask, why such a function? Well, this function will have many amazing properties that will simplify our computations later on. Of course, we could always abandon pgf's, but introducing this will be very useful as you will see.

Property 1. $G_X(1) = 1$.

This is because the sum of probabilities is always equal to 1.

Property 2. $G'_X(1) = E(X)$ where G'_X is the derivative of G_X.

This is because $G'_X(x) = \sum np_n x^{n-1}$ and plugging in $x = 1$ gives $G'_X(1) = \sum np_n = E(X)$.

Property 3. $G''_X(1) + G'_X(1) - G'_X(1)^2 = Var(X)$ where G''_X is the derivative of G'_X.

This is because $G''_X(1) + G'_X(1) = \sum n(n-1)p_n + \sum np_n = \sum n^2 p_n = E(X^2) = Var(X) + E(X)^2$.

Property 4. If X, Y are independent, $G_{X+Y}(x) = G_X(x) \cdot G_Y(x)$.

This is because $\mathcal{P}(X + Y = n) = \sum_{x+y=n} \mathcal{P}(X = x) \cdot \mathcal{P}(Y = y)$ which is the coefficient of x^n in $G_X \cdot G_Y$.

Property 5 (Convolution). If $X_1, \cdots, X_n$ are independent, then $G_{\sum X_i}(x) = \prod G_{X_i}(x)$.

This is just the generalisation of property 4, proven by induction as $X_1 + \cdots + X_n = (X_1 + \cdots + X_{n-1}) + X_n$.

Property 6. If a, b are integers, $G_{aX+b} = x^b \cdot G_X(x^a)$.

$X = n \Leftrightarrow aX + b = an + b$, so every x^n in $G_X(x)$ is replaced to x^{an+b} to get G_{ax+b}.

2 Binomial Distribution

Assume an experiment has exactly two outcomes, one with probability p, the other with $q = 1 - p$ (such an experiment is called a Bernoulli trial with probability of success p). If we repeat the experiment 3 times independently, the outcomes are SSS, SSF, SFS, SFF, FSS, FSF, FFS, FFF. The probability of getting 2 successes is

$$\begin{aligned}
&\mathcal{P}(SSF \text{ or } SFS \text{ or } FSS) \\
&= p \cdot p \cdot (1-p) + p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot p \\
&= 3p^2 q.
\end{aligned}$$

We can generalise this. Say we repeat the experiment n times independently, then there are $\binom{n}{k}$ combinations where exactly k successes appear. Each of these combinations have a probability of $p^k(1-p)^{n-k}$ because the n experiments are independent. Therefore the probability that there are k successes in a binomial distribution $\mathcal{B}(n,p)$ with n trials and success probability p is

$$B_{n,p}(k) = \binom{n}{k} p^k q^{n-k} \qquad \text{where } q = 1-p.$$

2.1 Properties

The pgf of $X \sim \mathcal{B}(n,p)$ is

$$G_X(x) = \sum_{k=0}^{n} B_{n,p}(k) x^k = \sum_{k=0}^{n} \binom{n}{k} q^{n-k} (px)^k = (q+px)^n.$$

Therefore

$$G'_X(x) = np(q+px)^{n-1} \qquad \text{and} \qquad G''_X(x) = n(n-1)p^2(q+px)^{n-2}$$

and thus $E(X) = np$ and $Var(X) = n(n-1)p^2 + np - n^2p^2 = npq$.

3 Geometric Distribution

Assume we have a Bernoulli trial again with probability of success p. Instead of finding the number of successes under n trials, suppose we are finding the number of trials needed to obtain the first success. The probability of getting $\underbrace{F \cdots F}_{k-1} S$ is $(1-p)^{k-1}p$, and thus if $X(\sim \text{Geo}(p))$ is the number of trials needed,

$$\mathcal{P}(X = k) = q^{k-1}p.$$

3.1 Properties

The pgf of $X \sim \text{Geo}(p)$ is

$$G_X(x) = \sum_{k=1}^{\infty} (1-p)^{k-1} p x^k = px \sum_{k=1}^{\infty} (x - px)^{k-1} = \frac{px}{1-qx}.$$

Therefore

$$G'_X(x) = \frac{p}{(1-qx)^2} \qquad \text{and} \qquad G''_X(x) = \frac{2pq}{(1-qx)^3}$$

and thus $E(X) = 1/p$ and $Var(x) = 2q/p^2 + 1/p - 1/p^2 = q/p^2$.

4 Poisson Distribution

We will now describe the Poisson[1] distribution. Imagine we have a situation where a Bernoulli trial is repeated a huge number of times, but the success probability is so small that only a few successes will happen. A notable example of this would be earthquakes. Earthquakes could potentially happen at any second, so every second there is a Bernoulli trial deciding whether an earthquake will happen, but in a year (which has 3×10^7 seconds), there will only be a few number of earthquake incidents, say 10 on average.

We could even subdivide a year into smaller intervals than seconds. But still, only 10 earthquakes happen per year on average. This is a binomial distribution $\mathcal{B}(n, p)$ where n is very large and $np = 10$ is constant. Therefore, it makes sense to let $n \to \infty$ and analyse that distribution instead. This will be called the *Poisson distribution* with parameter $\lambda = 10$.

Supposing $n \to \infty$, let's find the probability that there will be k earthquakes in a year (10 on average does not mean every year it is 10). Since $p = \lambda/n$, the probability is

$$\begin{aligned}
\lim_{n\to\infty} B_{n,p}(k) &= \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \lim_{n\to\infty} (1)(1)\cdots(1)\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^n (1)^{-k} \\
&= \frac{\lambda^k}{k!} e^{-\lambda}.
\end{aligned}$$

where we used the fact that $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$. In other words, we say X has the Poisson distribution $\mathrm{Po}(\lambda)$ with parameter λ if, for $k = 0, 1, \cdots$,

$$\mathcal{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In conclusion, the Poisson distribution is technically binomial with infinite trials and finite mean λ.

4.1 Properties

The pgf of $X \sim \mathrm{Po}(\lambda)$ is

$$G_X(x) = \sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda} = e^{\lambda x} \cdot e^{-\lambda} = e^{\lambda(x-1)}.$$

Therefore $E(X) = \lambda$ and $Var(x) = \lambda^2 + \lambda - \lambda^2 = \lambda$. Alternatively, we can see that $E(X) = \lim_{n\to\infty} np = \lim_{n\to\infty} \lambda = \lambda$ and $Var(X) = \lim_{n\to\infty} np(1-p) = \lim_{n\to\infty} \lambda(1-\lambda/n) = \lambda$.

[1] pronounced 'pwah-sawn'

5 Uniform Distribution

Fix a positive integer n. Assume $\mathcal{P}(X = k) = 1/n$ for $k = 1, \cdots, n$. Then X is said to have the uniform distribution $\mathcal{U}(n)$.

5.1 Properties

The pgf of $X \sim \mathcal{U}(n)$ is

$$G_X(x) = \sum_{k=1}^{n} \frac{x^k}{n} = \frac{x + x^2 + x^3 + \cdots + x^n}{n} = \frac{x(1 - x^n)}{n(1 - x)}$$

and also $E(X) = (1 + \cdots + n)/n = (n + 1)/2$ and $Var(X) = (1^2 + \cdots + n^2)/n = (n + 1)(2n + 1)/6$.

References

[1] Binomial Distribution
https://en.wikipedia.org/wiki/Binomial_distribution

[2] Geometric Distribution
https://en.wikipedia.org/wiki/Geometric_distribution

[3] Poisson Distribution
https://en.wikipedia.org/wiki/Poisson_distribution

[4] Uniform Distribution
https://en.wikipedia.org/wiki/Discrete_uniform_distribution

[5] T-Distribution
https://en.wikipedia.org/wiki/Student%27s_t-distribution