Sample Problems
Lecture 3: Probability Distributions and their Statistics

Problems prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

1. The Moment Generating Function

The moment generating function is a definition that usually facilitates the calculation of non-central moments for a probability distribution. This concept can be used both for discrete and continuous probability distributions. The moment generating function for a probability distribution denoted by its PDF, f_X (given the information I) is calculated as follows:

$$\mathbb{M}_X(t) = \int_{\Omega_X} e^{tx} f_X(x|\mathrm{I})\,\mathrm{d}x \equiv \mathbb{E}_X\left(e^{tx}|\mathrm{I}\right)$$

where t is a deterministic parameter.

- **Calculate the first derivative of $\mathbb{M}_X$ with respect to *t* at *t*=0.**

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{M}_X\big|_{t=0} = \int_{\Omega_X} xe^{tx} f_X(x|\mathrm{I})\,\mathrm{d}x\Big|_{t=0} = \int_{\Omega_X} x\, f_X(x|\mathrm{I})\,\mathrm{d}x \equiv \mathbb{E}\left(X|\mathrm{I}\right)$$

- **Calculate the *n*th derivative of $\mathbb{M}_X$ with respect to *t* at *t*=0.**

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathbb{M}_X\big|_{t=0} = \int_{\Omega_X} x^m e^{tx} f_X(x|\mathrm{I})\,\mathrm{d}x\Big|_{t=0} = \int_{\Omega_X} x^m\, f_X(x|\mathrm{I})\,\mathrm{d}x \equiv \mathbb{E}\left(X^m|\mathrm{I}\right)$$

- **Extend the result of the previous question to the case of discrete distribution.**

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathbb{M}_X\big|_{t=0} = \sum_x x^m e^{tx} P\left(X=x|\mathrm{I}\right)\Big|_{t=0} = \sum_x x^m P\left(X=x|\mathrm{I}\right) \equiv \mathbb{E}\left(X^m|\mathrm{I}\right)$$

2. First-excursion collapse probability assessment for a structure subjected to an aftershock sequence (Geometric Distribution)

A strong earthquake event is usually followed by a sequence of (often smaller-in-magnitude earthquake events called as the *Aftershocks*). The ensemble of these aftershocks is also referred to as a *Seismic Sequence*.

Suppose that the probability of structural collapse given that an earthquake of interest has happened is equal to π (assuming that each time an earthquake occurs, it hits the intact structure; a clearly un-realistic assumption!). Assume that in an interval of time $[0,t]$, n aftershock events have taken place.

- **What is the probability that the *k*th aftershock is going to lead to structural collapse (for the first time)?**

$A \equiv k$th aftershocks leads to structural collapse (kth trial is the first success)

$\mathbf{I} \equiv$ the probability of collapse (success) is π

Having (k-1) aftershocks (with no structural collapse) followed by one success (collapse), it is easily to verify that the PMF can be estimated as:

$$P\left(A|\mathbf{I}\right) = P\left(k|\pi\right) = \left(1-\pi\right)^{k-1}\pi$$

Note: The above distribution is called the *Geometric distribution*, which is the probability that the first occurrence of success requires k number of independent trials knowing that the probability of success of a single trial is equal to π, (i.e., <u>how many trials up to the first success</u>)

- **What is the moment generating function for *k*? (hint: $\mathbb{M}_K(t) = \sum_k e^{tk} P(k)$, you can use the formula for the sum of a geometric series).**

$$\begin{aligned}\mathbb{M}_K(t) &= \sum_{k=1}^{+\infty} e^{tk}(1-\pi)^{k-1}\pi = \frac{\pi}{1-\pi}\sum_{k=1}^{+\infty}\left[e^t(1-\pi)\right]^k \\ &= \frac{\pi}{1-\pi}e^t(1-\pi)\left[1+e^t(1-\pi)+\left[e^t(1-\pi)\right]^2+\cdots\right] \\ &= \pi e^t \frac{1}{1-e^t(1-\pi)}\end{aligned}$$

Note that we have used the limit of the sum of a Geometric series for $e^t(1-\pi)<1$.

- **What is the expected value for *k*? (hint: use the moment generating function).**

The first moment (expected value) can be calculated as (see Problem 1):

$$\mathbb{E}(K) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{M}_K\big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\pi e^t \frac{1}{1-e^t(1-\pi)}\right)\Big|_{t=0} = \frac{1}{\pi}$$

- **What is the best-estimate for *k* (e.g., expected value plus one standard deviation)?**

Accordingly, the second moment and the variance can be obtained from:

$$\mathbb{E}(K^2) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbb{M}_K\big|_{t=0} = \frac{(2-\pi)}{\pi^2}$$

$$\mathbb{VAR}(K|\mathrm{I}) = \mathbb{E}(K^2|\mathrm{I}) - \left[\mathbb{E}(K|\mathrm{I})\right]^2 = \frac{(2-\pi)}{\pi^2} - \frac{1}{\pi^2} = \frac{(1-\pi)}{\pi^2}$$

$$\mathit{Best\ Estimate} = \frac{1}{\pi} + \frac{\sqrt{1-\pi}}{\pi} = \frac{1+\sqrt{1-\pi}}{\pi}$$

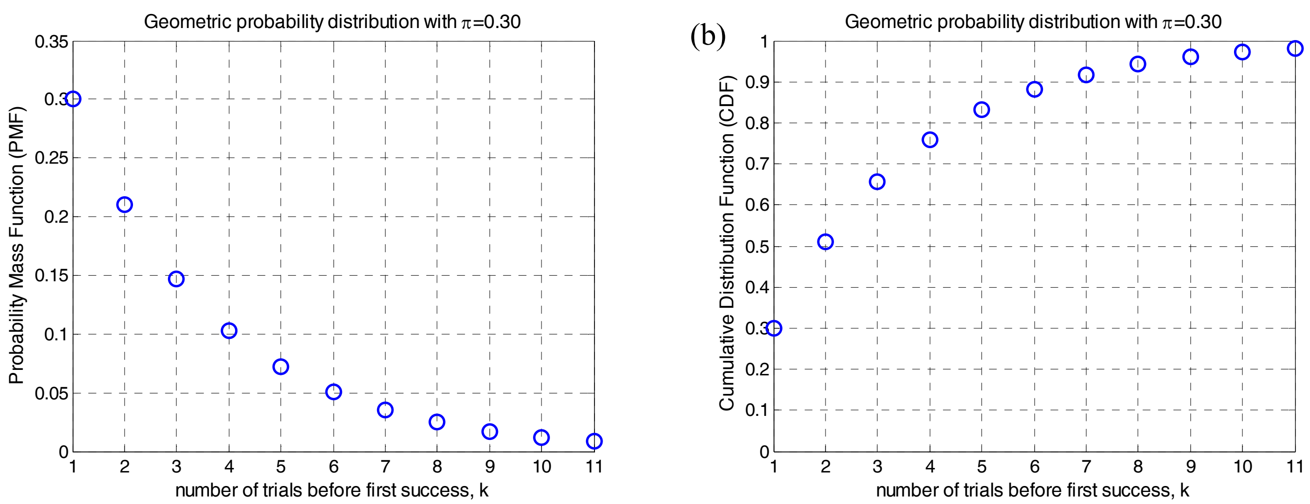
- **What is the probability that one of the first *n* aftershocks that hit the structure causes structural collapse for the first time, also called the first-excursion collapse limit state probability (hint: use the result obtained in part one and calculate the probability that *k* is smaller than or equal to *n*)?**

$$P\left(k \le n \middle| \pi\right) = P\left((k=1)+\cdots+(k=n) \middle| \pi\right) = \sum_{k=1}^{n} P\left(k \middle| \pi\right) = \pi \sum_{k=1}^{n} \left(1-\pi\right)^{k-1}$$

$$= \pi \sum_{i=0}^{n-1} \left(1-\pi\right)^{i} = \pi \frac{1-\left(1-\pi\right)^{n}}{1-\left(1-\pi\right)} = 1-\left(1-\pi\right)^{n}$$

Accordingly, this is actually the CDF of a Geometric distribution.

- **Plot the PMF and PDF for a Geometric distribution assuming that π=0.001.**



The geometric distribution (a) probability mass function (PMF), (b) cumulative distribution function (CDF)

3. Calculating the expected direct repair costs for the severe damage limit state (Negative Binomial Distribution)

The direct repair costs incurred to a given infrastructure for a given limit state (e.g., Severe Damage) are referred to as those costs that are incurred due to direct consequence of a strong earthquake for that limit state (i.e., severe damage).

Suppose that the probability that the infrastructure exceeds the severe damage limit state given that an earthquake of interest has happened is equal to π (assuming that each time the structure exceeds the severe damage limit state, it is repaired back to its intact state; looking at the entire life of an infrastructure, it is somewhat reasonable!). Moreover, assume that the cost of repair associated to the severe damage limit state is equal to C_D.

- **What is the probability distribution for the number of earthquakes *n* that need to happen in order for the number of repair operations to reach a given value *k*?**

$A \equiv$ *n*th events (trial) lead to *k*th repair (success)

$\mathbf{I} \equiv$ the occurrence of *n* event (success), and the probability of success of a single trial is π

Note: The problem actually introduce *Negative binomial* distribution, which shows how many trials should be performed up to the occurrence of the *k*th success. We have 1 success exactly at *n*th trial with probability π. Moreover, we have had (*k*-1) success out of (*n*-1) independent trials, which has a Binomial probability distribution as $P(k-1|n-1,\pi)$. Hence, the PMF can be estimated as:

$$P\left(A|\mathbf{I}\right)=P\left(n|k,\pi\right)=\pi\cdot P\left(k-1|n-1,\pi\right)=\pi\cdot\binom{n-1}{k-1}\pi^{k-1}\left(1-\pi\right)^{n-k}=\binom{n-1}{k-1}\pi^{k}\left(1-\pi\right)^{n-k}$$

- **What is the expected value for the number of earthquakes *n* that need to happen in order for the number of repair operations to reach a given value *k*? (hint: use the parameters of the Geometric distribution and the fact that we have the sum of *k* independent and identically distributed geometric variables)**

The Negative Binomial distribution is actually the sum of k **i**ndependent (and hence uncorrelated) and **i**dentically **d**istributed (*i.i.d*) geometric variables. Hence, the expected value will become:

$$\mathbb{E}\left(n \middle| k,\pi\right) = \sum_{i=1}^{k} \frac{1}{\pi} = \frac{k}{\pi}$$

- **What is the standard deviation for the number of earthquakes *n* that need to happen in order for the number of repair operations to reach a given value *k*?**

The variance will become:

$$\mathbb{VAR}\left(n \middle| k,\pi\right) = \sum_{i=1}^{k} \frac{\left(1-\pi\right)}{\pi^2} = \frac{k\left(1-\pi\right)}{\pi^2}$$

- **Now, imagine that exactly *n* earthquake events of interest have taken place in time interval [0,*t*]. What is the probability that the infrastructure has to be repaired exactly *k* times during interval [0,*t*]?**

$$P\left(k \middle| n,\pi\right) = \binom{n}{k} \pi^k \left(1-\pi\right)^{n-k}$$

- **What is the expected value for *k*?**

$$\mathbb{E}\left(k \middle| n,\pi\right) = \sum_{i=1}^{n} \pi = n\pi$$

- **What is the expected repair costs for the Severe Damage limit state?**

$$\mathbb{E}\left(\text{repair cost}\right) = n\pi C_D$$