Statistics Lecture

Claudio Ortelli [1]

[1]Finance Institute
USI

Advanced Learning and Research Institute, 2012

Bernoulli Trials

Consider a random experiment with only two outcomes, “success” and “failure”, and denote the probability of the two outcomes by p and q, respectively, with $p + q = 1$.

Definition

The experiment consisting in observing a sequence of n independent repetitions of the above described experiment is called a sequence of Bernoulli trials.

Examples

1. Observe n consecutive executions of an if statement, with success = “then clause is executed” and failure = ”else clause is executed”.
2. Examine components produced on an assembly line, with success = ”acceptable” and failure = ”defective”.

Bernoulli Trials

Let 0 denote failure and 1 denote success. Let S_n be the sample space of an experiment involving n Bernoulli trials

$$
\begin{aligned}
S_1 &= \{0,1\}, \\
S_2 &= \{(0,0),(0,1),(1,0),(1,1)\}, \\
S_n &= \{0,1\}^n = \{2^n\, n-\text{tuples of 0s and 1s}\}.
\end{aligned}
$$

For all sample spaces S_n we define the $\sigma-$ algebra $\mathscr{P}(S_n)$ as the relevant $\sigma-$ algebra on which to define the probability P. On S_1 we then have $P(\{0\}) = q$ and $P(\{1\}) = p$. We wish to assign probability to points in S_n.
Define $A_i =$"success on trail i" and $\bar{A}_i =$"failure on trial i". We then have $P(A_i) = p$ and $P(\bar{A}_i) = q$. Let s be an outcome of S_n with k "1" and $n-k$ "0", i.e.

$$s = (1,1,\ldots,1,0,0,\ldots 0)$$

Bernoulli Trials

The elementary event $\{s\}$ can be written

$$\{s\} = A_1 \cap A_2 \cap \cdots \cap A_k \cap \bar{A}_{k+1} \cap \cdots \cap \bar{A}_n.$$

Because events A_i are independent we obtain

$$P(\{s\}) = P(A_1)P(A_2)\ldots P(A_k)P(\bar{A}_{k+1})\ldots P(\bar{A}_n)$$

so that $P(\{s\}) = p^k q^{n-k}$. Note that that we can construct $\left(\begin{array}{c} n \\ k \end{array}\right)$ different outcomes with k successes and $n-k$ failures, therefore defining $A =$"we observe exactly k successes in n trials"

$$P(A) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k q^{n-k}.$$

Since by the binomial theorem $(p+q)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) p^k q^{n-k} = 1$, P is a well defined probability law on $(S_n, \mathscr{P}(S_n))$.

Bernoulli Trials

In connection with reliability theory let us assume that a particular system with n components requires at least k components to function in order for the entire system to work correctly. Such systems are called k-out-of-n systems.

- If we let $k = n$ we have a series system.
- If we let $k = 1$ we have a system with parallel redundancy.

Assuming all components are statistically identical and function independently of each other, and denoting by R the reliability of a component ($q = 1 - R$ gives its unreliability), then the experiment of observing the statuses of n components can be thought of as a sequence of n Bernoulli trials with probability $p = R$.

$$
\begin{aligned}
R_{k|n} &= P(\text{"at least } k \text{components functioning properly"}) \\
&= P(\bigcup_{i=k}^{n} \text{"exactly } i \text{components functioning properly"})
\end{aligned}
$$

$$
\begin{aligned}
R_{k|n} &= \sum_{i=k}^{n} P(\text{"exactly } i \text{components functioning properly"}) \\
&= \sum_{i=k}^{n} \binom{n}{i} p^i q^{n-i}.
\end{aligned}
$$

Example

Triple modular redundancy (TMR).

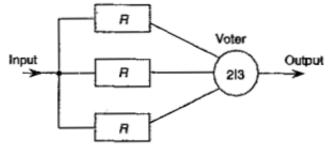


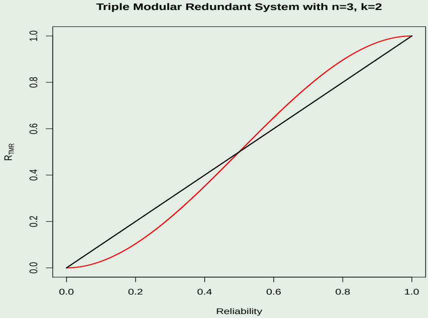
Figure 1.19. A triple modular redundant system

Bernoulli Trials

Example

Applying the above formula with $n = 3$ and $k = 2$:

$$R_{TMR} = \sum_{i=2}^{3} \binom{3}{i} R^i (1-R)^{3-i} = 3R^2(1-R) + R^3.$$



Non homogeneous Bernoulli Trials

When the components are nonhomegeneous w.r.t. the corresponding reliabilities, then the calculation is a bit more complicated:

$$R_{k|n} = 1 - \sum_{|I|>=k} \left(\prod_{i \in I} (1 - R_i) \right) \left(\prod_{i \notin I} R_i \right)$$

where I ranges over all choises $i_1 < i_2 < \cdots < i_m$ such that $k \leq m \leq n$ and R_i denotes the reliability of th $i-$th component.

Example

Consider a non homogeneous TMR with $n = 3$ and $k = 2$.

$$\begin{aligned} R_{2|3} &= 1 - (1 - R_1)(1 - R_2)R_3 - (1 - R_1)(1 - R_3)R_2 + \\ &\quad -(1 - R_2)(1 - R_3)R_1 - (1 - R_1)(1 - R_2)(1 - R_3) \\ &= R_1 R_2 + R_2 R_3 + R_1 R_3 - 2R_1 R_2 R_3 \end{aligned}$$

Generalized Bernoulli Trials

Next, we consider **generalized Bernoulli trials**. Here we have a sequence of n independent trials, and on each trial the result is exactly one of the k possibilities $b_1, b_2, \ldots, b_k$. On a given trial, let b_i occur with probability p_i, $i = 1, 2, \ldots, k$ such that

$$p_i \geq 0 \text{ and } \sum_{i=1}^{k} p_i = 1.$$

The sample space S consists of all k^n n-tuples with components $b_1, b_2, \ldots, b_k$. To a point $s \in S$

$$s = (\underbrace{b_1, b_1, \ldots, b_1}_{n_1}, \underbrace{b_2, b_2, \ldots, b_2}_{n_2}, \ldots, \underbrace{b_k, \ldots, b_k}_{n_k})$$

we assign the probability of $p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$, where $\sum_{i=1}^{k} n_i = n$. This is the probability assigned to any n-tuple having n_i occurrences of b_i, where $i = 1, 2, \ldots, k$. The number of such n-tuples are given by the multinomial coefficient [LIU 1968]:

$$\binom{n}{n_1 \;\; n_2 \;\; \cdots \;\; n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

As before, the probability that b_1 will occur n_1 times, b_2 will occur n_2 times, $\ldots$, and b_k will occur n_k times is given by

$$p(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \tag{1.23}$$

and

$$\begin{aligned} \sum_{n_i \geq 0} p(n_1, n_2, \ldots, n_k) &= (p_1 + p_2 + \cdots + p_k)^n \\ &= 1 \end{aligned}$$

Generalized Bernoulli Trials

Example

Trivedi, page 52.

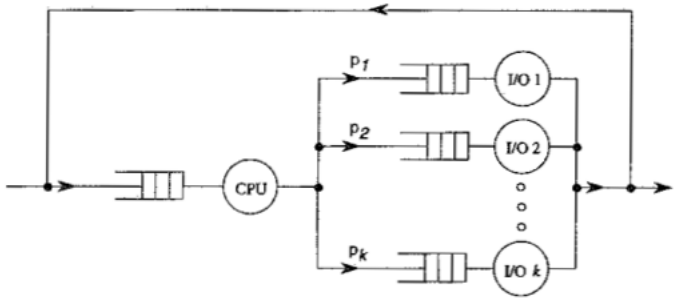


Figure 1.23. A CPU to I/O device queuing scheme

Solve problems 2,3,4 at page 56 and review problem 1 at page 57 of Trivedi.

Random variables

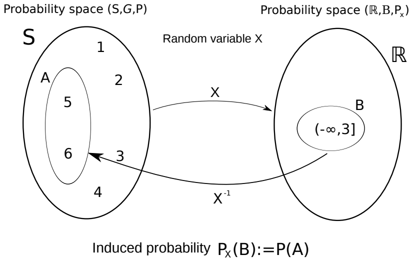
Let $(\Omega, \mathscr{G})$ denote a measurable space with sample space Ω and $\sigma-$algebra $\mathscr{G}$.

Definition

A random variable is a real valued function $X : \Omega \to \mathbb{R}$ such that

$$X^{-1}(B) \in \mathscr{G} \text{ for all events } B \in \mathscr{B}.$$

The following picture clarifies the situation



Random variables

Remarks:

1. A random variable is a function between two measurable spaces satisfying a *measurability* condition: the preimage of any event $B \subset \mathbb{R}$ must be an event of $\mathscr{G}$.
2. S can be finite, countable infinite or uncountable. If the image of X is discrete, i.e. finite or countable, then X is a discrete random variable.

Definition

Let X be a random variable and $x \in \mathbb{R}$ a real number. The event

$$A_x = \{s \in S \mid X(s) = x\}$$

is called the inverse image of the set $\{x\}$ and represents the outcomes of the random experiment which are mapped to x.

Random variables

It is clear that $A_x \cap A_y = \emptyset$ if $x \neq y$ and that

$$\bigcup_{x \in \mathbb{R}} A_x = S.$$

Attention:

1. Unions over an uncountable number of events are not, in general, events (see the definition of $\sigma-$algebra).
2. If the random variable X is discrete, then $\bigcup_{x \in I} A_x$ is an event for all $I \subset \mathbb{R}$. Why?

Notation:

1. The notation $[X = x]$ will be used as an abbreviation for A_x.
2. Similarly, $[X \leq x]$ denotes the event $E = \{s \in S \mid X(s) \leq x\}$. Analogous definitions apply for the other inequality operators.

Random variables

The set of random variables is closed under addition and scalar multiplication, under maximum, minimum, multiplication and division as well as under limit operation.

Theorem

Let X and Y two random variables defined on the same measurable space $(S, \mathscr{G})$. then

1. $aX + bY$ *is a random variable for all* $a, b \in \mathbb{R}$.
2. $\max\{X, Y\}$ *and* $\min\{X, Y\}$ *are random variables.*
3. XY *is random variable.*
4. *Provided that* $Y(s) \neq 0$ *for each* $s \in S$, X/Y *is a random variable.*
5. *Let* $X_1, X_2, \ldots$ *be a sequence of random variables. If*

$$X(s) = \lim_{n \to \infty} X_n(s)$$

exists for every $s \in S$, *the* X *is a random variable.*

Probability mass function

Let X be a discrete random variable and $I \subset \mathbb{R}$ the image set of X. Because X is discrete, set I is countable and

$$\sum_{x \in I} P(X = x) = \sum_{x \in I} P(A_x) = 1.$$

Furthermore, for all $x \notin I$, $P(X = x) = 0$ while for all events $B \in \mathscr{B}$ the event $A_B = [X \in B] = \{s \in S \mid X(s) \in B\}$ so that the probability

$$P(X \in B) = P(\bigcup_{x \in B} A_x) = P(\bigcup_{x \in B \cap I} A_x) = \sum_{x \in B \cap I} P(A_x).$$

The probability of any event $B \in \mathscr{B}$ can be computed as a sum over a countable number of points $x \in B \cap I$.

Definition

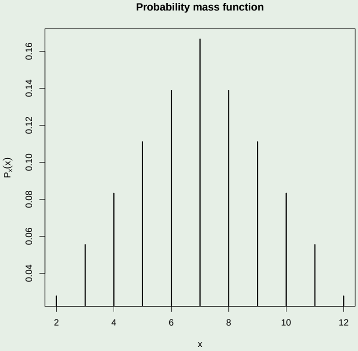
We define the *probability mass function* or *discrete density function* of the random variable X, denoted by p_X, the function

$$p_X(x) = P(A_x) \text{ for all } x \in \mathbb{R}.$$

Probability mass function

Example

A fair die is tossed twice. Let S denote the sample space of this random experiment and define X to be the sum of the outcomes of the first and second toss. The image of X is the set $I = \{2, 3, \ldots, 12\}$. The corresponding probability mass function is given in the following plot:

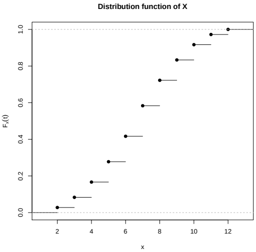


Cumulative distribution function

Definition

Let X be a discrete random variable and let $A_{(-\infty,t]}$ be the preimage of the event $(-\infty, t]$, $t \in \mathbb{R}$. The *cumulative distribution function* (CDF) or the probability distribution function or the *distribution function* of X is

$$\begin{aligned} F_X(t) &= P(X \leq t) = P(A_{(-\infty,t]}) \\ &= \sum_{x \in (-\infty,t] \cap I} P(A_x) = \sum_{x \in (-\infty,t] \cap I} p_X(x) \end{aligned}$$



Cumulative distribution function

A cumulative distribution F has the following properties:

1. $0 \leq F(x) \leq 1$ for $-\infty < x < \infty$. This follows because F is a probability.
2. F is a monotone increasing function of x, i.e. $F(x_1) \leq F(x_2)$ if $x_1 \leq x_2$ (think about the corresponding events ...).
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. If the random variable X has a finite image, then $F(x) = 0$ (1) for x sufficiently small (large).
4. Let $x_1, x_2, \ldots$ the elements of the image I of X. F has a positive jump equal to $p_X(x_i)$ at $i = 1, 2, \ldots$ and in the interval $[x_i, x_{i+1})$ F has a constant value $F(x_i)$.

It can be shown that any function F satisfying properties 1-4 is the distribution function of some discrete random variable!

Special discrete distributions

- The Bernoulli pmf is the density function of a discrete random variable X having 0 or 1 as its only possible values.
- The Binomial pmf is the density function of a discrete random variable Y_n which denotes the number of successes in n independent Bernoulli trials where each of them has a probability of success equal to p.
- The geometric pmf is the density function of a random variable Z describing the necessary number of independent Bernoulli trials in order to obtain the first success. The sample space of this random experiment is described by

$$S = \left\{ 0^{i-1}1 \mid i = 1, 2, 3, \dots \right\}.$$

  This sample space has an infinite number of outcomes.

Special discrete distributions

- The negative Binomial pmf is the density function of a random variable Z that describes the number of trials of a Bernoulli experiment which are necessary in order to obtain the $r^{th}-$success. It generalizes the geometric distribution.
- The Hypergeometric pmf is the density function of a random variable X wich computes the number k of defective components in a random sample of m components, chosen without replacement, from a total of n components, d of which are defective.
    - n, the number of components
    - d, the number of defective components
    - m, the number of components to sample (without replacement!)
    - k, the number of defective components found in the sample
- The uniform discrete pmf, the pmf of the constant random variable and the pmf of the indicator R.V.

Special discrete distributions

The Poisson distribution is the last discrete distribution we want to study. It is a very popular and very used distribution in probability and statistics because of its peculiarity having a pmf equal to

$$f(k;\alpha) = e^{-\alpha}\frac{\alpha^k}{k!} \text{ with } k = 0, 1, 2, \ldots$$

The Poisson distribution can be derived as a limit problem involving the binomial distribution. Suppose we want to observe the arrival of jobs to a large database server for the time interval $[0, t)$. It is reasonable to assume that for a small interval of duration $\triangle t$ the probability of a new job arrival is $\lambda \cdot \triangle t$, where λ is a constant depending on the population using the server. If $\triangle t$ is sufficiently small, the probability of observing two or more arrivals in an interval $\triangle t$ can be neglected. We are interested in calculating the probability of k jobs arriving in the interval $[0, t)$ under this hypothesis. It can be shown that the limit distribution of the problem is the Poisson distribution (see Trivedi, pages 82-83).

Exercises

Solve problems 2,3,4,5 at page 91 of Trivedi.

Probability generating function

The probability generating function is a convenient tool which simplifies calculations involving nonnegative, integer valued random variables.

Definition

The probability generating function (PGF) of a nonnegative integer valued random variable X is

$$G_X(z) = \sum_{i=0}^{\infty} p_i z^i$$

$G_X(z)$ is also known as the $z-$transform of X and converges for all $z \in \mathbb{C}$ with $\mid z \mid < 1$. By definition of G_X we have that $G_X(1) = 1$.

Theorem

If two discrete random variables X and Y have the same $PGFs$, then they have the same distributions and probability mass functions.

Probability generating function

Exercise

Compute the PGF of the following discrete distributions:

1. *The binomial random variable.*
2. *The Bernoulli random variable.*
3. *The Uniform $U[1, n]$ distribution.*
4. *The constant random variable $X = 3$.*
5. *The Poisson distribution.*
6. *Solve probems 1 and 2 at page 99 of Trivedi.*

Discrete random vector

Motivation: We are often interested in the relationship between two or more (discrete) random variables. Assume we have two softwares implementing the same functionality. We are interested in the execution time in seconds (rounded to the nearest integer). Let us denote by X_1 and X_2 the execution time of the first and second software, respectively. Examples of events on which we could be interested in are $\{X_1 > 4, X_2 > X_1\}$, $\{X_1 \leq 5, X_2 \leq 4\}$ and so on.

Definition

A random vector $X = (X_1, \ldots, X_r)$ is an $r-$dimensional vector-valued function $X : S \to \mathbb{R}^r$, $s \mapsto X(s) = (X_1(s), \ldots, X_r(s))$. A discrete $r-$dimentional random vector is a function from S to $\mathbb{R}^r$ taking on finite or countably infinite set of vector values x.

Discrete random vector

Definition

The compound (or joint) pmf for a random vector X is defined to be

$$\begin{aligned} p_X(x) &= P(X = x) \\ &= P(X_1 = x_1, \ldots, X_r = x_r). \end{aligned}$$

Fact

The joint pmf of a discrete random vector has the following 4 properties:

1. $p_X(x) \geq 0$ *for all* $x \in \mathbb{R}^r$*;*
2. *The set* $\{x \in \mathbb{R}^r \mid p_X(x) \geq 0\}$ *is a finite or countable infinite subset of* $\mathbb{R}^r$ *which is denoted by* $\{x_i,\ i = 1, 2, \ldots\}$ *;*
3. $P(X \in A) = \sum_{x \in A} p_X(x)$*;*
4. $\sum_{i=1}^{\infty} p_X(x_i) = 1.$

Every real valued function defined on $\mathbb{R}^r$ satisfying properties 1-4 is the joint pmf of some discrete $r-$dim. random vector.

Discrete random vector

Example

Consider a program with two modules with stochastic execution times X and Y, respectively. Assume for simplicity that the images of the discrete random variables X and Y are $\{1,2\}$ and $\{1,2,3,4\}$. The joint pmf of the random vector $Z = (X, Y)'$ is described by the following table:

	$y = 1$	$y = 2$	$y = 3$	$y = 4$
$X = 1$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$
$X = 2$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

The marginal pmf of the random variable X is by definition $p_X(x) = P(X = x)$ and can be derived from the joint pmf by "integrating out" the unrelevant random variables (Y in this case).

Discrete random vector

Example

(Continued).

$$
\begin{aligned}
p_X(x) &= P(X = X) = P(\bigcup_j \{X = x, Y = y_j\}) \\
&= \sum_j P(X = x, Y = y_j) \\
&= \sum_j p_Z(x, y_j).
\end{aligned}
$$

Examples

Let X and Y be two random variables, each with image $\{1,2\}$ and with a joint pmf given by

$$\begin{aligned} P_{X,Y}(1,1) &= P_{X,Y}(2,2) = a, \\ P_{X,Y}(1,2) &= P_{X,Y}(2,1) = 1/2 - a, \text{ for } 0 \leq a \leq \frac{1}{2}. \end{aligned}$$

It is easy to see that $p_X(1) = p_X(2) = p_Y(1) = p_Y(2) = \frac{1}{2}$, whatever be the value of a. This means we have uncountable many distinct joint pmf's associated with the same marginal pmf's!

Discrete random vector

Examples

The multinomial pmf. Consider a sequence of n generalized Bernoulli trials where at each trial there are the same finite number of distinct outcomes having probabilities $p_1, p_2, \ldots, p_r$ where $\sum_{i=1}^{r} p_i = 1$. Define the random vector $X = (X_1, X_2, \ldots, X_r)$ such that X_i is the number of trials that resulted in the $i-th$ outcome. The joint pmf of X is then given by

$$
\begin{aligned}
p_X(n) &= P(X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r) \\
&= \underbrace{\begin{pmatrix} n \\ n_1 n_2 \cdots n_r \end{pmatrix}}_{\frac{n!}{n_1! n_2! \cdots n_r!}} p_1^{n_1} p_1^{n_1} \cdots p_r^{n_r}
\end{aligned}
$$

See Trivedi, page 102 for a practical example involving the multinomial distribution!

Discrete random vector

Examples

An inspection plan calls for inspecting five chips and for either accepting each chip, rejecting each chip or submitting it for reinspection, with probabilities of $p_1 = 0.70$, $p_2 = 0.20$ and $p_3 = 0.10$, respectively. Questions:

1. What is the probability that all five chips must be reinspected?
2. What is the probability that none of the chips must be reinspected?
3. What is the probability that at least one of the chips must be reinspected?

Solve problems 1 and 2 at page 104 of Trivedi.

Definition

(Independent Random Variables). Two discrete random variables X and Y are defined to be independent if and only if their joint pmf is the product of their marginal pmf's, i.e.

$$P_{X,Y}(x,y) = p_X(x)p_Y(y) \text{ for all } x \text{ and } y.$$

Independent random variables

Let assume X and Y be two independent random variables. Define the random variable $Z = X + Y$. The event $\{Z = t\}$ can be represented by all the outcomes on the line $X + Y = t$.

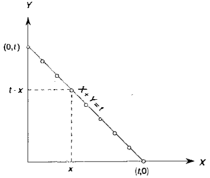


Figure 2.20. Computing the pmf of the random variable $Z_1 = X + Y$

$$
\begin{aligned}
P(Z = t) &= \sum_{x=0}^{t} P(X = x, Y = t - x) \\
&= \sum_{x=0}^{t} P(X = x)P(Y = t - x)
\end{aligned}
$$

Independent random variables

From previous formula it follows that

$$p_Z(t) \quad = \quad \sum_{x=0}^{t} p_X(x) p_Y(t-x).$$

This summation is called the discrete convolution and it gives the formula for the pmf of the sum of two nonnegative, indipendent, discrete random variables.

Definition

Let $X_1, X_2, \ldots, X_r$ be r discrete random variable with pmf's $p_{X_1}, p_{X_2}, \ldots, p_{X_r}$, respectively. The R.V. are said to be *mutually independent* if and only if their joint pmf is given by

$$p_{X_1, X_2, \ldots X_r}(x_1, x_2, \ldots x_r) \quad = \quad p_{x_1}(x_1) p_{x_2}(x_2) \cdots p_{x_r}(x_r).$$

Remark: it is possible for every pair of R.V. in the set $\{X_1, \ldots, X_r\}$ to be pairwise independent without the entire set to be mutually independent.

Independent random variables

Example

Consider a sequence of two Bernoulli trials and define X_1 and X_2 as the number of successes on the first and second trials, respectively. Let X_3 define the number of matches on the two trials. Then it can be shown that the pairs (X_1, X_2), (X_1, X_3) and (X_2, X_3) are each independent, but that the set $\{X_1, X_2, X_3\}$ is not mutually independent.

Restricting our attention to nonnegative integer-valued R.V. and recalling the definition of probability generating function PGF, the PGF of the sum of two independent random variables is the product of their PGFs:

$$\begin{aligned} G_Z(z) &= G_{X+Y}(z) \\ &= G_X(z)G_Y(z). \end{aligned}$$

Proof:

$$
\begin{aligned}
G_Z(z) &= \sum_{t=0}^{\infty} p_Z(t) z^t \\
&= \sum_{t=0}^{\infty} z^t \sum_{x=0}^{t} p_X(x) p_Y(t-x) \\
&= \sum_{t=0}^{\infty} \sum_{x=0}^{t} z^t p_X(x) p_Y(t-x) \\
&= \sum_{x=0}^{\infty} \sum_{t=x}^{\infty} z^t p_X(x) p_Y(t-x) \\
&= \sum_{x=0}^{\infty} z^x p_X(x) \sum_{t=x}^{\infty} z^{t-x} p_Y(t-x) \\
&= G_X(z) G_Y(z).
\end{aligned}
$$