MATH 5670-6670 Fall 2019

Transformed random variables

1 With a Poisson process

In BP_16 (Slides 7 and 13) we considered the sum of iid Bernoulli random variables stopped at N, a Poisson random variable. Actually, we may take a general value of a Poisson process with intensity λ. We will also rewrite the sum using a Boolean notation:

$$R_t = \sum_{i=1}^{N_t} X_i = \sum_i X_i \, \mathbb{1}_{\{i \le N_t\}}. \tag{1}$$

Next, we will use the duality

$$\{\, i \le N_t \,\} = \{\, S_i \le t \,\} \quad \Rightarrow \quad R(t) = \sum_i X_i \, \mathbb{1}_{\{S_i \le t\}}.$$

In lieu of the interval $[0, t]$ we may use any set A:

$$R(A) = \sum_i X_i \, \mathbb{1}_{\{S_i \in A\}}.$$

That is, whenever the i^{th} signal falls into a set A, X_i is added to the "collection".

It is called Poisson **reward process**, with X_n's as "rewards"[1].

When X_i are 01-Bernoulli RVs, the signal is kept with probability p and skipped with probability $1 - p$. The original Poisson process has been thinned attaining a new reduced intensity $p\lambda$. However, in general a reward Poisson process is not Poisson itself but may be viewed as an "abstract Poisson process". That is, the count of random points (S_n, X_n) falling into a bounded region in the plane has a Poisson distribution.

Exercise 1.

- *Compute* $\mathsf{E}\, R_t$ *and* $\mathsf{Var}(R_t)$.
- *Compute the mgf* $M(x) = \mathsf{E}\, e^{xR_t}$ *(the usual variable* t *must be changed to, e.g.,* x *because* t *already denotes time). Use the mgf to derive the mean and variance in an alternative way by differentiation.*
- *Stop the process at a random time* T*, independent of the process itself and of the rewards. Compute the expected reward acquired before time* T*, i.e.,* $\mathsf{E}\, R_T$*.*

[1] Jorge Luis Borges *The Lottery in Babylon*

2 Simple transformations of known distributions

2.1 Empirical standardization of normal

When the mean μ and the variance σ^2 are known, we perform the usual affine transformation

$$X \quad \mapsto \quad \widetilde{X} = \frac{X-\mu}{\sigma}.$$

When the parameters are unknown, we estimate them from a normal sample using the unbiased (and efficient) estimators $\overline{X}$ and S. So, the empirical standardization

$$X \quad \mapsto \quad \widetilde{X} = \frac{X-\overline{X}}{S}.$$

yields the pattern

$$T = \frac{Z}{\sqrt{Y}},$$

where $Z \sim N(0,1)$ and Y is $\Gamma(\alpha, 2)$ (here $\alpha = \frac{r}{2}$ for a suitable r, yielding a χ^2_r variable).

Exercise 2.

- *Find the pdf of T. Sketch it.*
- *What is the the largest n such that $\mathsf{E}\,|T|^n$ exists?*

It's the pdf of **T-distribution** (a.k.a. **Student's T distribution**).

2.2 Ratio of Gammas

Previously (BP_16, Slides 14-16) the ratio was based on the transformation $\frac{x}{x+y}$.

Let us now consider the direct ratio $\psi(x,y) = \frac{x}{y}$.

Exercise 3.

Let $X \sim \Gamma(\alpha, \theta = 1)$ and $Y \sim \Gamma(\beta, \theta = 1)$ be independent.

- *Find the pdf of $F = \dfrac{X}{Y}$.*
- *Sketch it.*
- *Discuss its asymptotics near 0 and at ∞ (compare to power functions).*

In practice, we deal with the special case of χ^2 distributions, $\alpha = \frac{r}{2}$, $\beta = \frac{s}{2}$.
Then we obtain so called the **F distribution with parameters r and s**.
It serves as a means of comparison of empirical (thus, ideal) variances of two independent samples.

3 Exponentiation

3.1 Finances or Physics

In finances, a fund is a basic financial instrument allowing consecutive investments. Then the gain (or loss) depends on fluctuations of the fund. A simple model is represented by sums of random variables (X_n). That is, X_n is the increment (it may have a negative sign) of the fund in the n^{th} cycle of time (say, a day, a month, a year, etc.). So, the value of the fund after n days is

$$S_n = C + X_1 + \ldots + X_n, \tag{2}$$

where $C > 0$ is the starting capital (an honest fund must not start with 0).

At the bottom of the n^{th} cycle an investor invests c_n units. He or she knows the performance of the fund only up to the previous $(n-1)^{\text{th}}$ day. So, the gain on the n^{h} day is simply $c_n X_n$ and the total gain after n days is

$$W_n = c_1 X_1 + \ldots + c_n X_n. \tag{3}$$

A physical interpretation is also available. An object moves in each cycle of time by a random distance $X_1, X_2, ...$ reaching the position (2) after n cycles. Consecutive forces $c_1, c_2, ...$ are applied in each cycle, resulting in the total work done (3) after n cycles.

3.2 Rate of Change

The simplest model involves iid summands, possibly symmetric (i.e., X_n and $-X_n$ would have the same distribution).

We observe some limitations of this model. First, the fund S_n must be positive, At the moment it hits 0, it ceases to exist (under reasonable and legal circumstances). Yet, in the symmetric case, the probability of doubling the initial capital, i.e., reaching the level $2C$ is equal to the probability of hitting 0.

Secondly, the most important is not the sheer size of the fun but its rate of growth (decrease is understood as negative growth), a.k.a. "interest" (when expressed as percentage). It's simply

$$I_n = \frac{S_n - S_{n-1}}{S_{n-1}} = \frac{X_n}{S_{n-1}}. \tag{4}$$

This rate should not depend on the size S_{n-1} of the fund. However, it does, except when the increments are deterministic constants. Yet, it stands to reason to require independence.

Therefore, an adjusted model may be proposed, with the initial value $a = Y_0 > 0$

$$Y_n = ae^{bS_n} > 0.$$

W.l.o.g. we put $a = 1$. The interest on the n^{th} day is independent of the value Y_{n-1}:

$$I_n = \frac{Y_n - Y_{n-1}}{Y_{n-1}} = e^{bX_{n-1}} - 1.$$

3.3 Lognormal

The transformed normal $Y = e^{bZ}$ is called **lognormal** (indeed, $\ln Y \sim N(0, b^2)$).

Exercise 4.

- *Find the lognormal pdf.*
- *Sketch it.*

  (It's a "skewed bell", resembling some Gamma, Weibull, or F-densities).
- *Discuss its asymptotics near 0 and at ∞ (compare to power functions).*

  It differs significantly from the aforementioned skewed bells. How?

3.3.1 Fair game

A sequence Y_n is called a **fair game**[2] if we expect no relative change when it progresses, formally:

$$\mathsf{E}\left[Y_n | Y_{n-1} = y\right] = y, \quad \text{for any value } y.$$

For example, the sum $Y_n = X_1 + \cdots + X_n$ of independent RVs of mean 0 is a fair game. Indeed,

$$\mathsf{E}\left[Y_n | Y_{n-1} = y\right] = \mathsf{E}\left[Y_{n-1} | Y_{n-1} = y\right] + \mathsf{E}\left[X_n | Y_{n-1} = y\right] = y + \mathsf{E}\, X_n = y.$$

($\mathsf{E}\left[X | Y = y\right] = \mathsf{E}\, X$ follows directly from the definition of the conditional density.)

Exercise 5.

- *Show that the log-linearly adjusted sequence $Y_n = \exp\left\{ bZ_n - \frac{b^2}{2} n \right\}$ is a fair game (Z_n's are iid N(0,1)).*
- *Find a deterministic sequence b_n such that $Y_n = \exp\left\{ W_n - b_n \right\}$ is a fair game.*

  (the scaling parameter is unnecessary because it may be included in coefficients c_n.)

[2] a.k.a. a "martingale"; a term of French origin, referring to a certain betting strategy in horse races

3.3.2 Continuous time

In the case of a process with continuous time such as a Poisson process or reward Poisson process the definition must be adjusted to earlier times rather to the "immediate previous" time. So, a fair game means

$$\mathsf{E}\left[Y_t | Y_s = y\right] = y, \quad \text{for any value } y \text{ and any earlier time } s < t.$$

By the same token, the condition on the rate of change (4) should be redefined:

$$I_{s \mapsto t} = \frac{Y_t - Y_s}{Y_s} \quad and \quad Y_s \quad \text{should be independent for all } s < t. \tag{5}$$

A poisson process N_t is not a fair game but after centering $Y_t = N_t - \lambda t$ it becomes a fair game. Even then it fails (5).

Exercise 6.

- *Given $c > 0$, for what number $a = a(c)$ the process $Y_t = \exp\{ cN_t - at \}$ does become a fair game?*
- *Same question for the reward Poisson process: $Y_t = \exp\{ R_t - at \}$*

  (here the scaling parameter c is obsolete because rewards X_i may be already scaled).