TRANSFOMATIONS AND CONDITIONING in HD

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

MATH 5670-6670 FALL 2019

PROBABILITY I

November 18, 2019

Outline

Simple conditioning

A bivariate function $w = \psi(x, y)$ applied to RVs X and Y with a joint distribution (pmf or pdf) $f(x, y)$ yields the transformation $W = \psi(X, Y)$.

The first question is to find the distribution of W.

When X and Y are discrete, so is W, and then

$$
\begin{aligned}
f_W(w) &= \mathrm{P}(\psi(X, Y) = w) \\
&= \sum_y \mathrm{P}(\psi(X, y) = w | Y = y)\, f_Y(y).
\end{aligned}
$$

Upon independence

If additionally X and Y are independent, then:

$$f_W(w) = \sum_y \mathrm{P}(\psi(X, y) = w)\, f_Y(y).$$

We may try the mgf:

$$\mathrm{E}\, e^{tW} = \sum_y \mathrm{E}\, e^{t\psi(X,y)}\, f_Y(y) = \sum_x \sum_y e^{t\psi(x,y)}\, f_X(x)\, f_Y(y).$$

In the general case, this is as far as we can go. We could only hope that the algebraic equation $\psi(x, y) = w$ could be solved for x or the exponential would take a simple form.

Example 1: The sum

Let $\psi(x,y) = x + y$. So, $W = X + Y$ and thus for independent summands

$$f_{X+Y}(w) = \sum_y f_X(w-y)\, f_Y(y), \quad M_{X+Y} = M_X \cdot M_Y.$$

The pmf is called the **convolution** of two pmf's. We write

$$f_{X+Y} = f_X * f_Y$$

It can be iterated over and over.

Examples of convolution

Mgf is more convenient only for specific discrete convolutions when the addition of independent summands preserves the category

- $\mathrm{bin}(n, p) * \mathrm{bin}(m, p) = \mathrm{bin}(n + m, p)$,
- $\mathrm{negbin}(r, p) * \mathrm{negbin}(s, p) = \mathrm{negbin}(r + s, p)$,
- $\mathrm{Poisson}(\lambda) * \mathrm{Poisson}(\mu) = \mathrm{Poisson}(\lambda + \mu)$.

For other distributions the direct definition is recommended.

Exercise. Find the pmf of $X + Y$, where both summands are independent uniform on $\{1, \dots m\}$ (may choose a small m, say, $m = 6$). Repeat for three summands.

Do you wish to continue for more summands?

Example 2

Exercise.

Let $X \sim \mathrm{bin}(n, p)$ and let an independent $N \sim \mathrm{Poisson}(\lambda)$.
What is the distribution of $\mathrm{bin}(N, p)$?

(i.e., the number of Bernoulli trials has been randomized according to Poisson distribution).

Use the direct approach (for the mgf approach see Slide 13)

Conditioning

There is no simple general continuous analogs of the formulas in Slides 3 and 4. We may try the cdf:

$$\mathrm{P}(\psi(X, Y) \le w) = \int_{\mathbb{R}} \mathrm{P}(\Phi(X, y) \le w | Y = y)\, f_Y(y)\, dy,$$

where independence may make it slightly easier:

$$\mathrm{P}(\psi(X, Y) \le w) = \int_{\mathbb{R}} \mathrm{P}(\Phi(X, y) \le w)\, f_Y(y)\, dy,$$

while hoping for solvability of the involved inequality and a hopefully helpful differentiation with respect to w to obtain the pdf.

When the inequality is solvable?

E.g., for transformations $\psi(x, y)$:

$$x + y, \quad x - y, \quad xy, \quad \frac{x}{y}, \quad \frac{x}{x+y}, \quad \dots$$

The first one, the sum, for independent summands again defines the **convolution**. In the case of the sum the mgf may be helpful if the addition of independent summands preserves the category:

- $\mathrm{Gamma}(\alpha, \theta) * \mathrm{Gamma}(\beta, \theta) = \mathrm{Gamma}(\lambda + \beta, \theta)$,
- $\mathrm{N}(\mu_1, \sigma_1^2) * \mathrm{N}(\mu_2, \sigma_2^2) = \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

Iterated expectation

Fubini's Theorem has been introduced in Calculus 3 as a natural means to compute multiple integrals as iterated integrals, computable one by one.

By the same token, for independent RVs X and Y

$$\mathsf{E}\,\psi(X,Y) = \mathsf{E}_X\Big(\mathsf{E}_Y\psi(X,Y)\Big) = \mathsf{E}_Y\Big(\mathsf{E}_X\psi(X,Y)\Big).$$

That is, within the inner bracket one RV is temporarily fixed and treated as a deterministic parameter.

Example 3: product of standard normal RVs

Let X, Y be independent N(0,1). Put $W = XY$. The mgf:

$$M(t) = \mathrm{E}\, e^{tW} = \mathrm{E}\, e^{tXY} = \mathrm{E}_X \Big(\mathrm{E}_Y e^{tXY} \Big) = \mathrm{E}\, e^{t^2 X^2/2}.$$

We know that $X^2 \sim \Gamma(\alpha = 1/2, \theta = 2)$.

Hence the latter is its mgf at $t^2/2$. That is,

$$M(t) = \frac{1}{\sqrt{1-t^2}} = \frac{1}{\sqrt{1-t}} \cdot \frac{1}{\sqrt{1+t}} = Ee^{t(S-S')},$$

where S, S' are independent half-signals related to a Poisson process of unit intensity.

That is

$$XY \text{ and } S - S' \text{ are equidistributed.}$$

Definition. If X' is an independent copy of X, then $Y = X - X'$ is called its **symmetrization**.

Symmetrization is not unique. An independent random sign $S = \pm 1$ with probability $1/2$ each also yields a symmetrization SX.

How they differ can be seen through the mgf's:

$$M(t)M(-t) \quad (\text{of } X - X') \quad \text{vs.} \quad \frac{M(t) + M(-t)}{2} \quad (\text{of } SX)$$

Example 2 revisited.

Let $X \sim \mathrm{bin}(n, p)$ with independently randomized number of trials. That is, $n \mapsto N \sim \mathrm{Poisson}(\lambda)$, to wit:

$$X = S_n = \sum_{i=1}^{n} X_i, \quad \text{where } X_i \text{ are iid Bernoulli with parameter } p.$$

So, $M(t) = \mathsf{E}\, e^{tS_N} = \mathsf{E}_N \left(\mathsf{E}_S e^{tS_N} \right) = \mathsf{E}_N \left(1 + p(e^t - 1)\right)^N$.
Put $s = \ln\left(1 + p(e^t - 1)\right)$, i.e., $1 + p(e^t - 1) = e^s$. Hence

$$M(t) = \mathsf{E}\, e^{sN} = e^{\lambda(e^s - 1)} = e^{p\lambda(e^t - 1)}$$

That is, $S_N \sim \mathrm{Poisson}(p\lambda)$.

Example 4. The ratio of waiting times.

In a Poisson process with unit intensity S_α is the moment of the α^{th} signal. Let us find the distribution of the ratio

$$R = \frac{S_\alpha}{S_{\alpha+\beta}} = \frac{S}{S+T},$$

where $S \sim \Gamma(\alpha, 1)$ and $T \sim \Gamma(\beta, 1)$ are independent.

The mgf approach is not much hopeful but the transformation $\psi(s,t) = \frac{s}{s+t}$ promises a solvable inequality.

Beta distribution

The Beta function is related to the Gamma function:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}, \quad \alpha > 0, \beta > 0.$$

We'll see that the ratio R has the **Beta density**:

$$B(\alpha, \beta)\, r^{\alpha-1}(1-r)^{\beta-1}, \quad 0 < r < 1.$$

Denote by F_α, f_α and F_β, f_β the corresponding cdf's and pdf's.

Calculations for Beta

Let $0 < r \le 1$.

$$F(r) = \mathrm{P}(X \le r) = \mathrm{P}(S \le \frac{rT}{1-r}) = \int_0^\infty F_\alpha\left(\frac{rt}{1-r}\right) f_\beta(t)\,dt.$$

Differentiating with respect to r, by the Chain Rule:

$$f(r) = \int_0^\infty \frac{t}{(1-r)^2}\, f_\alpha\left(\frac{rt}{1-r}\right) f_\beta(t)\,dt.$$

Using the explicit formulas for Gamma densities and a little algebra

$$f(r) = \frac{r^{\alpha-1}}{\Gamma(\alpha)\Gamma(\beta)(1-r)^{\alpha+1}} \int_0^\infty t^{\alpha+\beta-1}\, e^{-\frac{t}{1-r}}\,dt.$$

Finish and comments

The substitution $t = (1 - r)u$ yields the sought-for density.

The parameters α and β yield quite a versatile collection of shapes.

Consider just the left end point 0:

$0 < \alpha < 1$	-	a vertical asymptote;
$\alpha = 1$	-	a constant > 0;
$1 < \alpha < 2$	-	a sharp increase from 0;
$\alpha = 2$	-	a linear increase from 0;
$\alpha > 2$	-	a flat increase from 0;

Combined with the right end point, it covers $5 \times 5 = 25$ scenarios.

Semantics

In undergraduate probability courses the function $M(t) = \mathrm{E}\, e^{tX}$ is required, being named after just one of its features - to generate moments by differentiation.

Yet, quite often moments or mgf's do not exist.

We have seen other "jobs" of the mgf's. They

- identify the distributions;
- allow to see and handle their limits;
- capture transformations;
- simplify proofs (e.g., tangled convolutions vs. products);
- give numerous insights to the complicated realm of probability.

What instead?

In future endeavors you may expect to encounter

- Laplace transform, $L(t) = \mathrm{E}\, e^{-tX}$, suitable for nonnegative X;
- Fourier transform $\varphi(t) = \mathrm{E}\, e^{itX}$ that always exists.

They also generate moments as the simplest and easiest use.

Formulas are easily adjusted by replacing t by $-t$ or by it.

They both (and the mgf, too) stem from the Laplace transform of complex variable $L(z) = \mathrm{E}\, e^{zX}$, providing a powerful link to numerous mathematical areas such as Real and Complex Analysis.