Common distributions, MGFs and entropy

Sofia Olhede

EPFL

September 23, 2020

Moment Generating Functions III

- The mean, variance and moment generating function of $X \sim \mathrm{Bin}(n, p)$ are given by

$$\mathbb{E}(X) = np, \quad \mathbb{V}\mathrm{ar}(X) = np(1-p), \quad M_X(t) = (1 - p + pe^t)^n.$$

- If $X = \sum_{i=1}^n Y_i$ where $Y_i \overset{\text{iid}}{\sim} \mathrm{Bern}(p)$ then $X \sim \mathrm{Bin}(n, p)$.
- A random variable X is said to follow the Geometric distribution with parameter $p \in (0, 1)$ denoted $X \sim \mathrm{Geom}(p)$, if
  - $\mathcal{X} = \{0\} \cup \mathbb{N}$.
  - $f(x; p) = (1-p)^x p$.
- The mean, variance and moment generating of $X \sim \mathrm{Geom}(p)$ are given by

$$\mathbb{E}(X) = \frac{1-p}{p}, \quad \mathbb{V}\mathrm{ar}(X) = \frac{1-p}{p^2}, \quad M_X(t) = \frac{p}{1-(1-p)e^t},$$

- the latter for $t < -\log(1-p)$.

Moment Generating Functions IV

- Let $\{Y_i\}_{i\geq 1}$ be an infinite collection of random variables, where

$$Y_i \overset{\text{iid}}{\sim} \text{Bern}(p). \quad \text{Let} \quad T = \min\{k \in \mathbb{N} : \; Y_k = 1\} - 1$$

  Then $T \sim \text{Geom}(p)$.

- A random variable X is said to follow the Negative Binomial distribution with parameter $p \in (0,1)$ and $r > 0$, denoted $X \sim \text{NegBin}(r,p)$ if

  * $\mathcal{X} = \{0\} \cup \mathbb{N}$.
  * $f(x;p) = \binom{x+r-1}{x}(1-p)^x p^r$.

Moment Generating Functions V

- The mean, variance and moment generating function of $X \sim \mathrm{NegBin}(r, p)$ are given by

$$\mathbb{E}(X) = r\frac{1-p}{p}, \quad \mathbb{V}\mathrm{ar}(X) = r\frac{1-p}{p^2}, \quad M_X(t) = \frac{p^r}{(1-(1-p)e^t)^r},$$

- the latter for $t < -\log(1-p)$.
- If $X = \sum_{i=1}^{r} Y_i$ where $Y_i \sim \mathrm{Geom}(p)$ then $X \sim \mathrm{NegBin}(r, p)$.
- A random variable X is said to follow a Poisson distribution with parameter $\lambda > 0$ denoted $X \sim \mathrm{Poisson}(\lambda)$ if
  - $\mathcal{X} = \{0\} \cup \mathbb{N}$.
  - $f(x; \lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$.

Moment Generating Functions VI

- The mean, variance and moment generating function of $X \sim \mathrm{Poisson}(\lambda)$ are given by

$$\mathbb{E}(X) = \lambda, \quad \mathbb{V}\mathrm{ar}(X) = \lambda, \quad M_X(t) = \exp\{\lambda(e^t - 1)\}.$$

- Let $\{X_n\}_{n\geq 1}$ be a sequence of $\mathrm{Binom}(n, p_n)$ random variables such that $p_n = \lambda/n$ for some constant $\lambda > 0$. Then $f_{X_n} \overset{n\to\infty}{\to} f_y$ where $Y \sim \mathrm{Poisson}(\lambda)$.
- Let $X \sim \mathrm{Poisson}(\lambda)$ and let $Y \sim \mathrm{Poisson}(\mu)$ be independent. The conditional distribution of X given $X + Y = k$ is $\mathrm{Binom}(k, \frac{\lambda}{\lambda+\mu})$.
- A random vector $\boldsymbol{X}$ in $\mathbb{R}^k$ is said to follow the Multinomial distribution with parameters $n \in \mathbb{N}$ and $p = \begin{pmatrix} p_1, & \ldots & p_k \end{pmatrix} \in (0,1)^k$, such that $\sum_{i=1}^k p_i = 1$, denoted $\boldsymbol{X} \sim \mathrm{Multi}(n, p_1, \ldots, p_k)$ if

Moment Generating Functions VII

- Take

  * $\mathcal{X} = \{0, 1, \ldots, n\}^k$, and
  * $f(x_1, \ldots, x_k; n, \{p_i\}_{i=1}^k) =$
  $\frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k} \mathrm{I}\left\{ \sum_{i=1}^k x_i = n \right\}$.

- The mean, variance, covariance and moment generating function are:

$$\mathbb{E}(X_i) = np_i, \quad \mathbb{V}\mathrm{ar}(X_i) = np_i(1 - p_i), \quad \mathbb{C}\mathrm{ov}(X_i, X_j) = -np_i p_j,$$

$$M_{\boldsymbol{X}}(\mathrm{u}) = \left(\sum_{i=1}^k p_i e^{u_i} \right)^n.$$

- The multinomial generalizes the binomial distribution: n independent trials, with k possible outcomes.

Moment Generating Functions VIII

- Lemma (Poisson and Multinomial)
  If $X_i \sim \mathrm{Poisson}(\lambda_i)$, $i = 1, \ldots, k$ are independent, then the conditional distribution of $\mathbb{X} = \left(X_1, \ldots X_k\right)^T$ given $\sum_{i=1}^k X_i = n$ is $\mathrm{Multi}(n; p_1, \ldots, p_k)$ with

$$p_i = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_k}.$$

- A random variable X is said to follow the uniform distribution with parameters $-\infty < \theta_1 < \theta_2 < \infty$ denoted $X \sim \mathrm{Unif}(\theta_1, \theta_2)$ if

$$f_X(x; \theta) = \begin{cases} (\theta_2 - \theta_1)^{-1} & \text{if} \quad x \in (\theta_1, \theta_2) \\ 0 & \text{o/w} \end{cases}.$$

Moment Generating Functions IX

- The mean, variance and moment generating function of $X \sim \mathrm{Unif}(\theta_1, \theta_2)$ are given by

$$\mathbb{E}(X) = \frac{\theta_1 + \theta_2}{2}, \quad \mathbb{V}\mathrm{ar}(X) = \frac{(\theta_2 - \theta_1)^2}{12}, \quad M_X(t) = \frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)},$$

  $t \neq 0$. We also specify $M(0) = 1$.

- A random variable X is said to follow the exponential distribution with parameter $\lambda > 0$ denoted $X \sim \mathrm{Expl}(\lambda)$ if

$$f_X(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases}.$$

Moment Generating Functions X

- The mean, variance and moment generating function of $X \sim \mathrm{Exp}(\lambda)$ are given by

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad \mathbb{V}\mathrm{ar}(X) = \frac{1}{\lambda^2}, \quad M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

- If X and Y are independent exponential random variables with rates λ_1 and λ_2, then $Z = \min(X, Y)$ are also exponential with the rate $\lambda_1 + \lambda_2$.
- Lack of memory characterisation.

  * Let $X \sim \mathrm{Exp}(\lambda)$. Then $\Pr(X \geq x + t | X \geq t) = \Pr(X \geq x)$.
  * Conversely: if X is a random variable such that $\Pr(X \geq > 0) > 0$ and

$$\Pr(X > x + s | X > t) = \Pr(X > s), \quad \forall t, s \geq 0,$$

  then there exists a $\lambda > 0$ such that $X \sim \mathrm{Exp}(\lambda)$.

Moment Generating Functions XI

- A random variable X has a gammma distribution with parameters α and β (the shape and rate of the distribution respectively), written as $X \sim \mathrm{Gamma}(\alpha, \beta)$ if

$$f_X(x; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases}$$

- The mean, variance and moment generating function of $X \sim \mathrm{Gamma}(\alpha, \beta)$ are given by

$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \quad \mathbb{V}\mathrm{ar}(X) = \frac{\alpha}{\beta^2}, \quad M_X(t) = \left(\frac{\beta}{\beta - t}\right)^\alpha, \quad t < \beta.$$

Moment Generating Functions XII

- If $Y_1, \dots, Y_\alpha \overset{\text{iid}}{\sim} \text{Exp}(\beta)$ then $Y = Y_1 + \cdots + Y_\alpha \sim \text{Gamma}(\alpha, \beta)$ (also see the Erlang distribution)
- The special case of $X \sim \text{Gamma}(\frac{k}{2}, \frac{1}{2})$ is the chi-square distribution on k degrees of freedom written as χ_k^2.
- A random variable X follows the normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ denoted $X \sim N(\mu, \sigma^2)$ if

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad x \in \mathbb{R}.$$

- The mean, variance and moment generating function of $X \sim N(\mu, \sigma^2)$ are given by

$$\mathbb{E}(X) = \mu, \quad \mathbb{V}\text{ar}(X) = \sigma^2, \quad M_X(t) = \exp(t\mu + t^2\sigma^2/2).$$

- In the case of $Z \sim N(0,1)$ (standard normal density) we use $f_Z(z) = \varphi(z)$ and $F_Z(z) = \Phi(z)$.

Moment Generating Functions XIII

- Lemma: Let $X \sim N(\mu, \sigma^2)$ and assume $a \neq 0$. Then $aX + b \sim N(a\mu + b, a^2\sigma^2)$. Furthermore

  $$F_X(x) = \Phi(\frac{x - \mu}{\sigma}),$$

  where $\Phi()$ is the standard normal CDF or $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right) du$.
- Corollary: let $X_1, \ldots X_n$ be independent random variables and let $X_i \sim N(\mu_i, \sigma_i^2)$. Take S_n as the sum of the X_i. Then

  $$S_n \sim N(\sum_i \mu_i, \sum_i \sigma_i^2).$$

Moment Generating Functions XIII

- First note that the MGF of a Gaussian is

$$M_X(t) = \int_{\mathbb{R}} \frac{e^{tu}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(u-\mu)^2\right) du \tag{1}$$
$$= \exp(\mu t + \sigma^2 t^2/2). \tag{2}$$

We assume $X_1, \dots X_n$ are independent $X_i \sim N(\mu_i, \sigma_i^2)$, and so for $Y = \sum_i X_i$ is

$$\begin{aligned} M_Y(t) &= \mathbb{E}_Y\{e^{tY}\} \\ &= \mathbb{E}_{X_1,\dots X_n}\{e^{t\sum X_i}\} = \prod_i \mathbb{E}\{e^{tX_i}\} \\ &= \prod_i \exp(\mu_i t + \sigma_i^2 t^2/2) \\ &= \exp(\sum_i \mu_i t + \sum_i \sigma_i^2 t^2/2). \end{aligned}$$

Entropy etc

- The entropy is used to measure the disorder of a random variable.
- The entropy of a random variable X is defined as

$$H(X) = -\mathbb{E}\{\log f_X(X)\}$$
$$= \begin{cases} -\sum_{x\in\mathcal{X}} f_X(x)\log\{f_X(x)\} & \text{if} \quad X \quad \text{discrete} \\ -\int_{x\in\mathcal{X}} f_X(x)\log\{f_X(x)\}dx & \text{if} \quad X \quad \text{continuous} \end{cases}$$

- The entropy is a measure of intrinsic disorder or unpredictability of a random system. The entropy can be thought of as a measure of the uncertainty of the random variable X. One can think of this as the missing information: the larger the entropy the less we know about X.
- It is related to the variance, but is not equivalent to the variance.

Entropy etc

- It can be shown that when X is a discrete random variable then
  * $H(X) \geq 0$.
  * $H(g(X)) \leq H(X)$ for any deterministic function g.
- Entropy is expressed in the unit bits. If log is replaced by lg then the unit is nats.
- Can we then use entropy to compare distributions?

Entropy etc

- Let $p(x)$ and $q(x)$ be two probability density (probability mass) functions on $\mathbb{R}$. We define the Kullback-Leibler divergence or relative entropy of q with respect to p as

$$\mathrm{KL}(q||p) \equiv \int_{\mathbb{R}} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx. \tag{3}$$

- By Jensen's inequality for $X \sim p(\cdot)$ we have

$$\mathrm{KL}(q||p) = \mathbb{E}_p\left\{-\log\frac{q(X)}{p(X)}\right\} \geq -\log\mathbb{E}_p\left\{\frac{q(X)}{p(X)}\right\} = 0, \tag{4}$$

  as q is unit norm.

- $p = q \Leftrightarrow \mathrm{KL}(q||p) = 0.$
- The KL divergence is not a metric as it both lacks symmetry and violates the triangle inequality, though symmetrized versions exist.