Technical notes on probability: The continuous case

Yang Xu

Probability theory is not restricted to describing discrete variables. This note summarizes some common facts about probabilities concerning continuous variables with an emphasis in the Gaussian distribution (Gauss, 1809).

1. If a random variable X has a discrete distribution, the probability function $f(\cdot)$ that maps its values to real numbers between 0 and 1 is known as the probability mass function (PMF), such that $f(x) = p(X = x)$. If a random variable is continuous in its values, the probability function is known instead as the probability density function (PDF),

$$\int_{x_0}^{x_1} f(x) = p(x_0 \leq X \leq x_1), \tag{1}$$

where $\int_x f(x)dx = 1$ $(-\infty \leq x \leq \infty)$. In this case, $f(x)$ may have a value greater than 1.

Example A
A uniform PDF for a continuous variable X specifies a distribution over the numerical range $[l, u]$ where l and u are the parameters that specify the minimal and maximal values of X. The probability density of X in this case is constant $f(x|l,u) = \frac{1}{u-l}$ and 0 elsewhere. If $u - l < 1$, e.g. $u = 0$ and $l = 0.5$, then it follows that $f(x|l,u) = 2 > 1$.

2. The cumulative distribution function (CDF) $F(\cdot)$ is a monotonically increasing function such that $F(x_0) = p(x \leq x_0)$. In the case of discrete variables, $F(x_0) = \sum_{x=-\infty}^{x_0} p(x)$. In the case of continuous variables,

$$F(x_0) = \int_{-\infty}^{x_0} f(x)dx. \tag{2}$$

It follows that $p(x > x_0) = 1 - F(x_0)$.

Example B
The CDF of a uniform distribution as in Example A is piece-wise linear in x such that $F(x) = 0$ for $x < l$, $F(x) = \frac{x-l}{u-l}$ for $l \leq x < u$, and $F(x) = 1$ for $x \geq u$.

3. The expectation or mean of a variable X with PDF $f(x)$ is:

$$E[X] = \int_x xf(x)dx. \tag{3}$$

Example C
The mean of a uniform distribution as in Example A is $E[X] = \int_l^u x\frac{1}{u-l}dx + \int_{x \in [u,l]} 0dx = \frac{1}{2}\frac{u^2-l^2}{u-l} + 0 = \frac{u+l}{2}$.

4. The variance of a variable X with PDF $f(x)$ is:

$$Var(X) = E[(X - E[X])^2] = \int_x (x - E[X])^2 f(x)dx. \tag{4}$$

Work this out yourself and show $Var(X)$ for a uniform distribution as in Example A is $\frac{(u-l)^2}{12}$.

5. A commonly used PDF is the Gaussian distribution, also known as the normal distribution, with the form:

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{5}$$

where μ is the mean (center), and σ is the standard deviation (variance$^{\frac{1}{2}}$, an indicator of average spread from the mean).

6. The area under a Gaussian PDF for the range $[\mu + 2\sigma, \mu - 2\sigma]$ covers about 95% of the total probability densities.

7. A standard normal distribution centers at 0 and has a standard deviation of 1:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \tag{6}$$

where Z takes the transformed value of X: $z = \frac{x-\mu}{\sigma}$, hence $z = 2$ and $z = -2$ corresponds to two standard deviations above and below the mean respectively.

8. Multiplication of two Gaussians renders a Gaussian distribution off by a normalizing constant:

$$\mathcal{N}(x|\mu_u, \sigma_u)\mathcal{N}(x|\mu_v, \sigma_v) \propto \mathcal{N}(x|\mu_w, \sigma_w), \tag{7}$$

where

$$\mu_w = \frac{\mu_u\sigma_v^2 + \mu_v\sigma_u^2}{\sigma_u^2 + \sigma_v^2}; \sigma_w^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma_u^2 + \sigma_v^2}. \tag{8}$$

Namely, the combined mean is a linear combination of (and hence lies between) the two individual means:

$$\mu_w = w_u\mu_u + w_v\mu_v, \text{ where } w_u = \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2} \text{ and } w_v = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}, \tag{9}$$

and the combined variance is no greater than either of the individual variances:

$$\sigma_w^2 = \frac{\sigma_v^2}{1 + \frac{\sigma_v^2}{\sigma_u^2}} \leq \sigma_v^2 \text{ or } = \frac{\sigma_u^2}{\frac{\sigma_u^2}{\sigma_v^2} + 1} \leq \sigma_u^2. \tag{10}$$

Example C (exercise)
Imagine that you watched your friend throwing a dart and the dart landed at horizontal position x on a dart board. You know for a fact that the dart throw tends to center at position μ with a constant variance (e.g. σ^2). Based on your prior belief, you also think that the expected horizontal position μ itself centers around position p with uncertainty (i.e. variance) s^2. Given these two pieces of information, you should be able to derive an analytical expression for the posterior probability of the expected position μ, by making use of the Bayes rule (in formulating this problem) and the property of Gaussian multiplication (in deriving an analytical expression for the posterior).

9. The Gaussian PDF of a multivariate variable in d dimensions is:

$$f(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \tag{11}$$

where $\boldsymbol{x}$ and $\boldsymbol{\mu}$ are d-dimensional vectors, and $\boldsymbol{\Sigma}$ is a $d \times d$ covariance matrix.

Example D
For a bivariate normal ($d = 2$), $\boldsymbol{x} = [x_1\ x_2]^t$, $\boldsymbol{\mu} = [\mu_1\ \mu_2]^t$, $\boldsymbol{\Sigma}_{11} = \sigma_1^2$, $\boldsymbol{\Sigma}_{22} = \sigma_2^2$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} = E[(x_1 - \mu_1)(x_2 - \mu_2)]$, $|\boldsymbol{\Sigma}| = \boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}$ is the determinant of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Sigma}^{-1} = \frac{1}{|\boldsymbol{\Sigma}|}\boldsymbol{\Sigma}'$ is the inverse of $\boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma}'_{11} = \boldsymbol{\Sigma}_{22}$, $\boldsymbol{\Sigma}'_{22} = \boldsymbol{\Sigma}_{11}$, and $\boldsymbol{\Sigma}'_{12} = \boldsymbol{\Sigma}'_{21} = -\boldsymbol{\Sigma}_{12}$.[1] How would you interpret the four values in $\boldsymbol{\Sigma}$ in a 2-dimensional space? (See if you have an intuition for the estimated dart position for Example C imagining now that it lies in a 2-dimensional space.)

10*. If two (sets of) variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly Gaussian (e.g. $[\boldsymbol{x}\ \boldsymbol{y}]^t \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$), then the conditional probability of one variable over another (e.g. $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{y}|\boldsymbol{x})$) is also Gaussian. Similarly, the marginal probability of each variable is also Gaussian (e.g. $f(\boldsymbol{x}) = \int_{\boldsymbol{y}} f(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{y}$ or $f(\boldsymbol{y}) = \int_{\boldsymbol{x}} f(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}$).

[1] A reference for matrix algebra can be found here: https://www.math.uwaterloo.ca/ hwolkowi/matrixcookbook.pdf.