Lecture 8. Conditional Distributions - introduction to Bayesian Inference[1]

Igor Rychlik

Chalmers
Department of Mathematical Sciences

Probability, Statistics and Risk, MVE300 • Chalmers • April 2010

[1] Bayesian statistics is a general methodology to analyse and draw conclusions from data.

The conditional cdf $P(X \leq x|Y = y)$. and pdf

Suppose that we observed the value of Y, e.g. we know that $Y = y$, but X is not observed yet. An important question is if the uncertainty about X is affected by our knowledge that $Y = y$, i.e. if

$$F(x|y) = \mathrm{P}(X \leq x|Y = y)$$

depends on y[2].

For continuous r.v. X, Y it is not obvious how to define conditional probabilities given that "$Y = y$", since $\mathrm{P}(Y = y) = 0$ for all y. As before we can intuitively reason that we wish to condition on "$Y \approx y$" then the conditional pdf of X given $Y = y$ is define by

$$f(x|y) = \frac{f(x, y)}{f(y)}, \quad F(x|y) = \int_{-\infty}^{x} f(\widetilde{x}|y)\,\mathrm{d}\widetilde{x}$$

is the conditional distribution.

---

[2]If X and Y are independent then obviously $F(x|y) = F_X(x)$ and Y gives us no knowledge about X.

Law of Total Probability

Let $A_1, \ldots, A_n$ be a partition of the sample space.

Then for any event B

$$\mathrm{P}(B) = \mathrm{P}(B|A_1)\mathrm{P}(A_1) + \mathrm{P}(B|A_2)\mathrm{P}(A_2) + \cdots + \mathrm{P}(B|A_n)\mathrm{P}(A_n)$$

If X and Y have joint density $f(x, y)$ and B is a statement about X, then

$$\mathrm{P}(B) = \int_{-\infty}^{+\infty} \mathrm{P}(B|Y = y) f_Y(y)\,\mathrm{d}y. \quad \mathrm{P}(B \mid Y = y) = \int_B f(x|y)\,\mathrm{d}x,$$

Bayes formula: In many examples the new piece of information is formulated in form of a statement that is true. For example C ="the wire passed preloading test of 1000kg", i.e. $C =$" $Y > 1000$" is true. If the likelihood $L(y) = \mathrm{P}(C|Y = y)$ is known then the density $f(y|C)$ is computed using Bayes formula $f(y|C) = c\mathrm{P}(C|Y = y)f(y)$.

Typical problems in safety of existing structure:

Suppose a wire has known strength y. Let X_1 be the maximal load during the first year of exploitation. Compute

$$\mathrm{P}(\text{"wire survives first years load"}) = \mathrm{P}(X_1 < y) = F_{X_1}(y).$$

In reality strength y is not known, r.v. Y models the uncertainty and

$$P_{\text{safe}} = \mathrm{P}(\text{"wire survives first years load"}) = \mathrm{P}(X_1 < Y).$$

Problems:

(a) How to compute probability $P_{\text{safe}} = \mathrm{P}(B)$, where $B = \text{"}X1 < Y\text{"}$?

(b) Suppose $B = X_1 < Y$ is true, what is the probability

$$P_{\text{safe}} = \mathrm{P}(\text{"wire survives second year load"}|B) = \mathrm{P}(X_2 < Y|B)?$$

(c) What is distribution of strength Y after surviving the first year load, i.e. $F_Y(y|B) = \mathrm{P}(Y \leq y|B)$?

Bayesian methods in risk evaluation - example:

In the following we shall be mostly interested in studying uncertainties in estimation of probabilities in the following setup. The "initiation" events A are defined and their concurrences are modeled by Poisson point process with intensity λ_A. In order for A to develop to an accident or catastrophe, some other unfortunate circumstances, described by event B, have to take place (B is called a "scenario"). For example, if A is "fire ignition" B could be "failure of sprinkler system".

Sometimes one needs multi-scenario event B, i.e. $B = B_1 \cup B_2$ where B_1, B_2, are excluding. The important parameters are λ_A, $p_1 = \mathrm{P}(B_1)$ and $p_2 = \mathrm{P}(B_2)$.

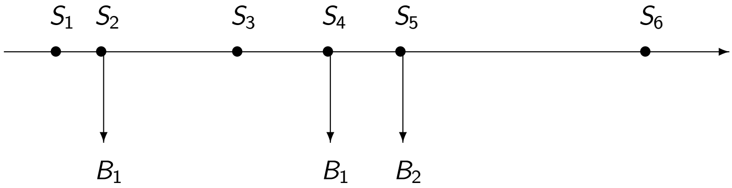

Figure: Events A at times S_i with related scenarios B_i.

$$P_t = P(\text{no accident in period } t) = 1 - \mathrm{e}^{-\lambda_A P(B)\, t} \approx \lambda_A P(B)\, t,$$

if probability P_t is small. Hence Two problems of interest in risk analysis:

- The first one will deal with the estimation of a probability $p_B = \mathrm{P}(B)$, say, of some event B, for example the probability of failure of some system.
- The second one is estimation of the probability that at least once an event A occurs in a time period of length t. The problem reduces itself to estimation of the intensity λ_A of A.

,

In general parameters p_B and λ_A are attributes of some physical system, *e.g.* if $B =$"A water sample passes tests" then $p_B = \mathrm{P}(B)$ is a measure of efficiency of a waste-water cleaning process. The intensity λ_A of accidents may characterize a particular road crossing. The parameters p_B and λ_A are unknown.

Odds for parameters

Let θ denote the unknown value of p_B, λ_A or any other quantity.

Introduce odds q_θ, which for any pair θ_1, θ_2 represents our belief which of θ_1 or θ_2 is more likely to be the unknown value of θ, *i.e.* $q_{\theta_1} : q_{\theta_2}$ are odds for the alternatives $A_1 =$ "$\theta = \theta_1$" against $A_2 =$ "$\theta = \theta_2$".

Since there are here uncountable number of alternatives, we require that q_θ integrates to one and hence $f(\theta) = q_\theta$ is a probability density function representing our belief about the value of θ.

Prior odds - posterior ods

Again, let θ be the unknown parameter, for example $\theta = p_B$, $\theta = \lambda_A$, while Θ denotes any of the variables P or Λ. Since θ is unknown, it is seen as a value taken by a random variable Θ with pdf $f(\theta)$.

If $f(\theta)$ is chosen on basis of experience without including observations of outcomes of an experiment then the density $f(\theta)$ is called a *prior density* and denoted by $f^{\text{prior}}(\theta)$.

However, as time passes, our knowledge may change, especially if we observe some outcomes of the experiment which can influence our opinions about the values of parameter θ reflecting in the new density $f(\theta)$. The modified density $f(\theta)$ will be called the *posterior density* and denoted by $f^{\text{post}}(\theta)$.

The method to update $f(\theta)$ is

$$f^{\text{post}}(\theta) = cL(\theta)\, f^{\text{prior}}(\theta)$$

How to find likelihood function $L(\theta)$ will be discussed later on.

Predictive probability

Suppose $f(p)$ has been selected and denote by P a random variable having pdf $f(p)$. A plot of $f(p)$ is an illustrative measure of how likely the different values of p_B are.

If only one value of the probability is needed, the Bayesian methodology proposes to use the so-called **predictive probability** which is simply the mean of P:

$$\mathsf{P}^{\text{pred}}(B) = \mathsf{E}[P] = \int p f(p)\,\mathrm{d}p.$$

The predictive probability measures the likelihood that B occurs in future. It combines two sources of uncertainty: the unpredictability whether B will be true in a future accident and the uncertainty in the value of probability p_B.

Example 6.1

Predictive probability

As before, if only one single value of the probability is needed, the Bayesian approach proposes to use the predictive probability

$$\begin{aligned} \mathsf{P}_t^{\text{pred}}(A) &= \mathsf{E}[P] = \int (1-\exp(-\lambda\, t)) f_\Lambda(\lambda)\,\mathrm{d}\lambda \\ &\approx \int t\lambda f_\Lambda(\lambda)\,\mathrm{d}\lambda = t\mathsf{E}[\Lambda].^3 \end{aligned}$$

This is a measure of the risk that A occurs, combining two sources of uncertainty: the variability of the Poisson process of A and the uncertainty in the intensity of accidents λ_A.

In some situations A is an initiation event (accident at a crossing) while B is scenario, e.g. $B =$"Victim needs hospitalisation". The intensity of $A \cap B$ is $\lambda_A \mathsf{P}(B)$. Uncertainty of $\lambda_A \mathsf{P}(B)$ is modeled by $\Lambda \cdot P$. The predictive probability of no serious accident is

$$\begin{aligned} \mathsf{P}_t^{\text{pred}}(A \cap B) &= \int (1-\exp(-p\lambda\, t)) f_\Lambda(\lambda) f_P(p)\,\mathrm{d}\lambda\,\mathrm{d}p \\ &\approx \int t\, p\lambda f_\Lambda(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}p = t\mathsf{E}[\Lambda]\mathsf{E}[P]. \end{aligned}$$

Example 6.2