BAYESIAN MODELLING OF PROFITABLE LANDING PAGE

Yana Bondarenko

Oles Honchar Dnipro National University, Ukraine
e-mail: yana.bondarenko@pm.me

Abstract

Each company applies great efforts on finding ways to make a profit. An effective website is important for the company because a significant part of the target audience receives information about the products and services via the Internet. Company website must be constantly optimized to bring in a profit. A/B testing is an experiment of showing two versions of the landing page to website visitors at the same time and comparing which one performs better for a given conversion aim. Our goal is to compute likelihood that the average income in version B is greater than the average income in version A.

Keywords: Bayesian inference, landing page optimization, likelihood, probability density function.

Let us assume that Bernoulli trials with probabilities of success λ_A, λ_B are conducted in groups of visitors in two versions of landing page. Probabilities of success λ_A, λ_B are unknown random variables. Note that success is a purchase. Suppose $p(\lambda_A)$, $p(\lambda_B)$ are the prior probability density functions for λ_A, λ_B, then $p(\lambda_A \mid x_1,...,x_n)$, $p(\lambda_B \mid y_1,...,y_m)$ are the posterior probability density functions for λ_A, λ_B. Let us remark that sample vectors $x=(x_1,...,x_n)$, $y=(y_1,...,y_m)$ are observed in two versions of landing page during the experiment.

In addition, suppose $f(\theta_A)$, $f(\theta_B)$ are the prior probability density functions for parameters θ_A, θ_B. Note also that these parameters θ_A, θ_B are some functions of the average size of purchase. Then $f(\theta_A \mid z_1,...,z_k)$, $f(\theta_B \mid w_1,...,w_l)$ are the posterior probability density functions for θ_A, θ_B. We stress that sample vectors $z=(z_1,...,z_k)$, $w=(w_1,...,w_l)$ are observed in two versions during the experiment.

Posterior distribution of probability of purchase in version A. Bernoulli trials with two possible outcomes (success is purchase, failure is no purchase) are conducted in version A. The number of successes (that is the number of purchases) in one trial has Bernoulli distribution with parameter λ_A (probability of purchase):

$$P(x,\lambda_A)=\lambda_A^x(1-\lambda_A)^{1-x},\ x=0,1;\ 0<\lambda_A<1. \tag{1}$$

The likelihood function is determined by:

$$p(x_1,\dots,x_n|\lambda_A)=\lambda_A^{x_1}(1-\lambda_A)^{1-x_1}\dots\lambda_A^{x_n}(1-\lambda_A)^{1-x_n}=\lambda_A^{\sum_{i=1}^{n}x_i}(1-\lambda_A)^{n-\sum_{i=1}^{n}x_i}, \tag{2}$$

where each x_i is equal to 1 or 0. The prior information about probability of purchase λ_A is defined by Beta distribution with parameters $a=1$, $b=1$ (equivalently, Uniform prior distribution as is generally known):

$$p(\lambda_A)=\frac{\lambda_A^{a-1}(1-\lambda_A)^{b-1}}{B(a,b)},\ 0\le\lambda_A\le 1. \tag{3}$$

According to Bayes' theorem, the posterior distribution for probability of purchase λ_A is given by:

$$p\left(\lambda_A | x_1,\ldots,x_n\right) = \frac{p\left(\lambda_A\right) p\left(x_1,\ldots,x_n|\lambda_A\right)}{\int_0^1 p\left(\lambda_A\right) p\left(x_1,\ldots,x_n|\lambda_A\right) d\lambda_A}. \tag{4}$$

Substituting $p\left(\lambda_A\right), p\left(x_1,\ldots,x_n|\lambda_A\right)$ in (4), we get:

$$p\left(\lambda_A | x_1,\ldots,x_n\right) = \frac{\lambda_A^{a-1}\left(1-\lambda_A\right)^{b-1} \lambda_A^{\sum_{i=1}^n x_i} \left(1-\lambda_A\right)^{n-\sum_{i=1}^n x_i}}{\int_0^1 \lambda_A^{a-1}\left(1-\lambda_A\right)^{b-1} \lambda_A^{\sum_{i=1}^n x_i} \left(1-\lambda_A\right)^{n-\sum_{i=1}^n x_i} d\lambda_A}. \tag{5}$$

Integrating (5) in λ_A, we obtain:

$$p\left(\lambda_A | x_1,\ldots,x_n\right) = \frac{\lambda_A^{a+\sum_{i=1}^n x_i -1} \left(1-\lambda_A\right)^{b+n-\sum_{i=1}^n x_i -1}}{B\left(a+\sum_{i=1}^n x_i, b+n-\sum_{i=1}^n x_i\right)}. \tag{6}$$

Thus, we have that the posterior distribution for probability of purchase λ_A is Beta distribution with parameters $(\tilde{a},\tilde{b})$:

$$p\left(\lambda_A | x_1,\ldots,x_n\right) = \frac{\lambda_A^{\tilde{a}-1}\left(1-\lambda_A\right)^{\tilde{b}-1}}{B\left(\tilde{a},\tilde{b}\right)}, \quad \tilde{a} = a + \sum_{i=1}^n x_i, \quad \tilde{b} = b + n - \sum_{i=1}^n x_i, \tag{7}$$

where n is the number of visitors and $\sum_{i=1}^n x_i$ is the number of complete purchases in version A.

Posterior distribution of probability of purchase in version B. Bernoulli trials with two possible outcomes (success, failure) are conducted in version B. The number of successes (number of purchases) in one trial has Bernoulli distribution with parameter λ_B (probability of purchase):

$$P\left(x,\lambda_B\right) = \lambda_B^y \left(1-\lambda_B\right)^{1-y}, \ y = 0, 1; \ 0 < \lambda_B < 1. \tag{8}$$

The likelihood function is determined by:

$$p\left(y_1, \ldots, y_m | \lambda_B\right) = \lambda_B^{y_1}\left(1-\lambda_B\right)^{1-y_1} \ldots \lambda_B^{y_m}\left(1-\lambda_B\right)^{1-y_m} = \lambda_B^{\sum_{i=1}^m y_i}\left(1-\lambda_B\right)^{m-\sum_{i=1}^m y_i}, \tag{9}$$

where each y_i is equal to 1 or 0. The prior information about probability of purchase λ_B is defined by Beta distribution with parameters $c = 1$, $d = 1$ (Uniform prior distribution):

$$p\left(\lambda_B\right) = \frac{\lambda_B^{c-1}\left(1-\lambda_B\right)^{d-1}}{B\left(c,d\right)}, \ 0 \le \lambda_B \le 1. \tag{10}$$

According to Bayes' theorem, the posterior distribution for probability of purchase λ_B is given by:

$$p\left(\lambda_B | y_1,\ldots,y_m\right) = \frac{p\left(\lambda_B\right) p\left(y_1,\ldots,y_m|\lambda_B\right)}{\int_0^1 p\left(\lambda_B\right) p\left(y_1,\ldots,y_m|\lambda_B\right) d\lambda_B}. \tag{11}$$

Using (9), (10), we get:

$$p\left(\lambda_B | y_1,\ldots,y_m\right) = \frac{\lambda_B^{c-1}\left(1-\lambda_B\right)^{d-1} \lambda_B^{\sum_{i=1}^m y_i} \left(1-\lambda_B\right)^{m-\sum_{i=1}^m y_i}}{\int_0^1 \lambda_B^{c-1}\left(1-\lambda_B\right)^{d-1} \lambda_B^{\sum_{i=1}^m y_i} \left(1-\lambda_B\right)^{m-\sum_{i=1}^m y_i} d\lambda_B}. \tag{12}$$

Finally, we obtain:

$$p\left(\lambda_B | y_1,\ldots,y_m\right) = \frac{\lambda_B^{c+\sum_{i=1}^m y_i -1} \left(1-\lambda_B\right)^{d+m-\sum_{i=1}^m y_i -1}}{B\left(c+\sum_{i=1}^m y_i, d+m-\sum_{i=1}^m y_i\right)}. \tag{13}$$

This yields that the posterior distribution for probability of purchase λ_B is Beta distribution with parameters $(\tilde{c},\tilde{d})$

$$p\left(\lambda_B|y_1,\ldots,y_m\right)=\frac{\lambda_B^{\tilde{c}-1}\left(1-\lambda_B\right)^{\tilde{d}-1}}{B\left(\tilde{c},\tilde{d}\right)},\quad \tilde{c}=c+\sum_{i=1}^{m}y_i,\quad \tilde{d}=d+m-\sum_{i=1}^{m}y_i, \tag{14}$$

where m is the number of visitors and $\sum_{i=1}^{m}y_i$ is the number of complete purchases in version B.

Posterior distribution of the size of purchase in version A. Now assume that the size of purchase has exponential distribution with parameter θ_A (that is reciprocal of the average size of purchase):

$$p\left(z,\theta_A\right)=\theta_A e^{-\theta_A z},\ z\geq 0,\ \theta_A>0. \tag{15}$$

The likelihood function is determined by:

$$p\left(z_1,\ \ldots,z_k|\theta_A\right)=\theta_A e^{-\theta_A z_1}\ldots\theta_A e^{-\theta_A z_k}=\left(\theta_A\right)^k e^{-\theta_A\sum_{i=1}^{k}z_i},\quad z_i>0\ \text{ for all }\ i\ . \tag{16}$$

The prior information about parameter θ_A is defined by Gamma distribution with parameters (s,t):

$$f\left(\theta_A\right)=\frac{t^s}{\Gamma\left(s\right)}\theta_A^{s-1}e^{-t\theta_A},\ \theta_A\geq 0,\ s>0,\ t>0. \tag{17}$$

According to Bayes' theorem, the posterior distribution for parameter θ_A is given by:

$$f\left(\theta_A|z_1,\ldots,z_k\right)=\frac{f\left(\theta_A\right)p\left(z_1,\ldots,z_k|\theta_A\right)}{\int_0^{+\infty}f\left(\theta_A\right)p\left(z_1,\ldots,z_k|\theta_A\right)d\theta_A}. \tag{18}$$

Substituting $f\left(\theta_A\right),p\left(z_1,\ldots,z_k|\theta_A\right)$ in (18), we get:

$$f\left(\theta_A|z_1,\ldots,z_k\right)=\frac{\theta_A^{s-1}e^{-t\theta_A}\left(\theta_A\right)^k e^{-\theta_A\sum_{i=1}^{k}z_i}}{\int_0^{+\infty}\theta_A^{s-1}e^{-t\theta_A}\left(\theta_A\right)^k e^{-\theta_A\sum_{i=1}^{k}z_i}d\theta_A}. \tag{19}$$

Integrating (19) in θ_A, we obtain:

$$f\left(\theta_A|z_1,\ldots,z_k\right)=\frac{\left(t+\sum_{i=1}^{k}z_i\right)^{s+k}}{\Gamma\left(s+k\right)}\theta_A^{s+k-1}e^{-\left(t+\sum_{i=1}^{k}z_i\right)\theta_A}. \tag{20}$$

Therefore, the posterior distribution for parameter θ_A is Gamma distribution with parameters $(\tilde{s},\tilde{t})$:

$$f\left(\theta_A|z_1,\ldots,z_k\right)=\frac{\tilde{t}^{\tilde{s}}}{\Gamma\left(\tilde{s}\right)}\theta_A^{\tilde{s}-1}e^{-\tilde{t}\theta_A},\ \tilde{s}=s+k,\ \tilde{t}=t+\sum_{i=1}^{k}z_i, \tag{21}$$

where k is the number of purchases and $\sum_{i=1}^{k}z_i$ is the sum of complete purchases in version A.

Posterior distribution of the size of purchase in version B. Suppose the size of purchase has exponential distribution with parameter θ_B (that is reciprocal of the average size of purchase):

$$p\left(w,\theta_B\right)=\theta_B e^{-\theta_B w},\ w\geq 0,\ \theta_B>0. \tag{22}$$

The likelihood function is determined by:

$$p\left(w_1,\ \ldots,w_l|\theta_B\right)=\theta_B e^{-\theta_B w_1}\ldots\theta_B e^{-\theta_B w_k}=\left(\theta_B\right)^l e^{-\theta_B\sum_{i=1}^{l}w_i},\quad w_i>0\ \text{ for all }\ i\ . \tag{23}$$

The prior information about parameter θ_B is defined by Gamma distribution with parameters (u,v):

$$f\left(\theta_b\right)=\frac{v^u}{\Gamma\left(u\right)}\theta_B^{u-1}e^{-v\theta_B},\ \theta_B\geq 0,u>0,v>0\ . \tag{24}$$

According to Bayes' theorem, the posterior distribution for parameter θ_B is given by:

$$f\left(\theta_B|w_1,\dots,w_l\right)=\frac{f\left(\theta_B\right)p\left(w_1,\dots,w_l|\theta_B\right)}{\int_0^{+\infty}f\left(\theta_B\right)p\left(w_1,\dots,w_l|\theta_B\right)d\theta_B}. \tag{25}$$

Using (22), (23), we get:

$$f\left(\theta_B|w_1,\dots,w_l\right)=\frac{\theta_B^{u-1}e^{-v\theta_B}\left(\theta_B\right)^l e^{-\theta_B\sum_{i=1}^l w_i}}{\int_0^{+\infty}\theta_B^{u-1}e^{-v\theta_B}\left(\theta_B\right)^l e^{-\theta_B\sum_{i=1}^l w_i}d\theta_B}. \tag{26}$$

Finally, we obtain:

$$f\left(\theta_B|w_1,\dots,w_l\right)=\frac{\left(v+\sum_{i=1}^l w_i\right)^{u+l}}{\Gamma\left(u+l\right)}\theta_B^{u+l-1}e^{-\left(v+\sum_{i=1}^l w_i\right)\theta_B}. \tag{27}$$

Hence, the posterior distribution for parameter θ_B is Gamma distribution with parameters $(\tilde{u},\tilde{v})$:

$$f\left(\theta_B|w_1,\dots,w_l\right)=\frac{\tilde{v}^{\tilde{u}}}{\Gamma\left(\tilde{u}\right)}\theta_B^{\tilde{u}-1}e^{-\tilde{v}\theta_B},\ \tilde{u}=u+l,\ \tilde{v}=v+\sum_{i=1}^l w_i, \tag{28}$$

where l is number of purchases and $\sum_{i=1}^l w_i$ is sum of accomplished purchases in landing page B.

Likelihood. The foregoing results allows us to describe probability of purchase λ and reciprocal of θ It is clear that λ_A is *the average number of complete purchases* in one trial in version A, λ_B is *the average number of complete purchases* in one trial in version B. Continuing in the same way, we see that $1/\theta_A$ is *the average size of complete purchases* in version A, $1/\theta_B$ is *the average size of complete purchases* in version B. Finally, *the average income* in version A is equal to λ_A/θ_A and *the average income* in version B is equal to λ_B/θ_B. Likelihood that the average income in version B is greater than the average income in version A is defined by:

$$P\left\{\frac{\lambda_B}{\theta_B}>\frac{\lambda_A}{\theta_A}\right\}=P\left\{\frac{\lambda_B}{\theta_B}-\frac{\lambda_A}{\theta_A}>0\right\}=$$

$$=\int_{\frac{\lambda_B}{\theta_B}-\frac{\lambda_A}{\theta_A}>0}p\left(\lambda_A,\theta_A,\lambda_B,\theta_B\mid x_1,\dots,x_n;z_1,\dots,z_k;y_1,\dots,y_m;w_1,\dots,w_l\right)d\lambda_A d\theta_A d\lambda_B d\theta_B=$$

$$=\int_{\frac{\lambda_B}{\theta_B}-\frac{\lambda_A}{\theta_A}>0}p(\lambda_A\mid x_1,\dots,x_n)f(\theta_A\mid z_1,\dots,z_k)p(\lambda_B\mid y_1,\dots,y_m)f(\theta_B\mid w_1,\dots,w_l)d\lambda_A d\theta_A d\lambda_B d\theta_B=$$

$$=\int_{\frac{\lambda_B}{\theta_B}-\frac{\lambda_A}{\theta_A}>0}\frac{\lambda_A^{\tilde{a}-1}\left(1-\lambda_A\right)^{\tilde{b}-1}}{B\left(\tilde{a},\tilde{b}\right)}\frac{\tilde{t}^{\tilde{s}}}{\Gamma\left(\tilde{s}\right)}\theta_A^{\tilde{s}-1}e^{-\tilde{t}\theta_A}\frac{\lambda_B^{\tilde{c}-1}\left(1-\lambda_B\right)^{\tilde{d}-1}}{B\left(\tilde{c},\tilde{d}\right)}\frac{\tilde{v}^{\tilde{u}}}{\Gamma\left(\tilde{u}\right)}\theta_B^{\tilde{u}-1}e^{-\tilde{v}\theta_B}d\lambda_A d\theta_A d\lambda_B d\theta_B.$$

Likelihood can be approximated with Monte Carlo sampling. Many computer languages have procedures for simulating this sampling process.

References

Bondarenko, Ya. (2019) Sequential A/B Testing. In: *Proceedings of the 2019 IEEE International Conference on Advanced Trends in Information Theory*, 488-491. https://ieeexplore.ieee.org/document/9030448

Bondarenko, Ya., Kravchenko, S. (2019) Bayesian approach to landing page testing. *Problems of Applied Mathematics and Mathematical Modelling*, 3-16. https://pm-mm.dp.ua/index.php/pmmm/article/view/240

Bondarenko, Ya. (2021) A sequential probability ratio test for randomized online experiments. *Problems of Applied Mathematics and Mathematical Modelling*, 3-15. https://pm-mm.dp.ua/index.php/pmmm/article/view/304

Stucchio, C. (2015) Bayesian A/B Testing at VWO. Whitepaper, Visual Website Optimizer.