Бајесовско оцењивање

За разлику од класичног (фреквенционистичког приступа), претпостављамо да су непознати параметри случајне величине које имају неку априорну расподелу. Природно се поставља питање коју? Могућности су разне и то пре свега зависи од нашег предзнања о неком параметру. Природно је да, када се добије нека нова информација, та расподела промени.

Основни мото Бајесовског приступа је да су вероватноће онолике колико ВЕРУЈЕМО да јесу.

Претпоставимо да предавач уђе први пут у учионицу на почетку семестра и постави неко питање које се односи на разумевање првог предавања. Означимо са p вероватноћу тачног одговора. Уколико предавач нема никакво предзнање о студентима јасно је да највише има смисла претпоставити да p има $\mathcal{U}[0,1]$ расподелу. Међутим како одмиче семестар, предавач постаје свестан састава публике и неће више претпостављати да се ради униформној расподелу. Дакле, априорна расподела се модификује. Фреквенционисти би "поновили први час" велики број пута и на основу тога оценили вероватноћу успеха. У пракси је то свакако немогуће јер се састав групе сваке године мења а и први час се дешава једном годишње.

Ради бољег разумевања навешћемо још један пример.

Пример 0.0.1. *Претпоставимо да се новчић баца $N = 3$ пута и да је од тих N пута $k = 3$ пута пало писмо. Желимо да оценимо вероватноћу да падне писмо. Уколико бисмо желели класихно да приступимо проблему и искористимо метод максималне веродостојности добили бисмо да је $\hat{p} = 1$.*

Наиме функција веродостојности је

$$L(p) \propto p^{\sum_{i=1}^{N} I\{x_i = pismo\}}(1-p)^{N - \sum_{i=1}^{N} I\{x_i = pismo\}} = p^3.$$

Максимум функције на интервалу $[0, 1]$ је баш за $p = 1$.

Када не знамо реализован узорак $\hat{p} = \bar{X}_n$.

Сада када бисмо поново бацили новчић очекивали бисмо да поново падне писмо. Што се не функционише баш тако у пракси.

Много је смисленије да претпоставимо да p има $\mathcal{U}[0, 1]$ расподелу и онда нађемо апостериорну расподелу за p уколико знамо да је било баш 3 писма.

$$
\begin{aligned}
f(p|\sum_{i=1}^{N} I\{x_i = P\} = k) &= \frac{P\{\sum_{i=1}^{N} I\{x_i = P\} = k|p\}\pi(p)}{\int_0^1 P\{\sum_{i=1}^{N} I\{x_i = P\} = k|p\}\pi(p)dp} \\
&= \frac{\binom{N}{k}p^k(1-p)^{N-k}\cdot 1}{\int_0^1 \binom{N}{k}p^k(1-p)^{N-k}\cdot 1dp}
\end{aligned} \tag{1}
$$

Случајна величина X има бета $\beta(\alpha, \beta)$ расподелу уколико је њена густина

$$
f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \;\; x \in [0,1], \;\; \alpha > 0, \; \beta > 0
$$

где је $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Тада је $E(X) = \frac{\alpha}{\alpha+\beta}$, *и* $D(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Одавде је

$$
\begin{aligned}
f(p|\sum_{i=1}^{N} I\{x_i = P\} = k) &= \frac{\binom{N}{k}p^k(1-p)^{N-k}\cdot 1}{\int_0^1 \binom{N}{k}p\cdot p^{k-1}(1-p)^{N-k}\cdot 1dp} \\
&= \frac{\binom{N}{k}p^k(1-p)^{N-k}\cdot 1}{E(\beta(k, N-k+1))} \\
&= \frac{\binom{N}{k}p^k(1-p)^{N-k}\cdot 1}{\frac{k}{N+1}} = Cp^k(1-p)^{N-k}.
\end{aligned}
$$

Дакле апостериорна расподела је $\beta(k+1, N-k+1)$.

Видимо да је класа расподела остала иста али да су се параметри променили. Сада се природно поставља питање шта је оцена за p. Има смисла да оцена буде баш мода апостериорне расподеле. У том случају добијамо исто сто и методом маскималне веродостојности. Међутим, како је некако природно мислити на просек када описујемо случајне величине, природно је да оцена буде баш очекивана вредност апостериорне расподеле.

Тада је $\hat{p} = \frac{k+1}{N+2} = \frac{4}{5}$.

Видимо да кад је N велико ова оцена и оцена методом максималне веродостојности се не разликују много.

Међутим, шта ако имамо неко предзнање о новчићу, нпр. да је више вероватно да је $p > 0.5$ *и да нпр. претпоставимо да је априорна* *распodела за p бета* $\beta(a,b)$*, за* $a = 3$ $b = 1$*. Тада се аналогним поступком може добити да је апостериорна расподела за p* $\beta(k+a, N-k+b)$ *па је* $\hat{p} = \frac{k+a}{N+a+b} = \frac{6}{7}$.

Видимо да оцена прилично зависи од априорне расподеле. Зато се о избору исте мора водити рачуна. Препорука је да када заиста немамо никаквог предзнања о параметру да користимо неку *неинформативну* расподелу.

Осим што зависи од априорне расподеле, оцена зависи и од тога коју нумеричку карактеристику апостериорне расподеле смо баш одабрали (имали смо пример моде и пример очекиване вредности). Свако ће природно бирати ону оцену која је по њему најбоља односно она која максимизира "корист", односно минимизира губитак. Управо зато морамо пре оцењивања да дефинишемо функцију губитака.

Неке од најчешћих функција губитака су

$$L(\hat{\theta}(X), \theta) = (\theta - \hat{\theta}(X))^2 \tag{2}$$

$$L(\hat{\theta}(X), \theta) = |\theta - \hat{\theta}(X)| \tag{3}$$

$$L(\hat{\theta}(X), \theta) = 1, \text{ за } |\theta - \hat{\theta}(X)| < \varepsilon \tag{4}$$

Одговарајући очекивани губици су

$$U(\hat{\theta}, \theta) = E(\theta - \hat{\theta}(X))^2 \tag{5}$$

$$U(\hat{\theta}, \theta) = E|\theta - \hat{\theta}(X)| \tag{6}$$

$$U(\hat{\theta}, \theta) = P\{|\theta - \hat{\theta}(X)| < \varepsilon\}. \tag{7}$$

Сада можемо дефинисати Бајесов ризик

$$R(\hat{\theta}) = \int_{-\infty}^{\infty} U(\hat{\theta, \theta}) f(\theta|x) d\theta. \tag{8}$$

Зато ћемо за Бајесову оцену узети баш ону вредност која минимизира Бајесов ризик. У случају (2) добија се да је оцена за θ $E(\theta|X)$, а у случају (3) $med(\theta|X)$. У случају (4), кад $\varepsilon \to 0$ добија се да је оцена за θ $argmax f(\theta|X)$.

С обзиром на природу парамтера θ можемо на основу апостериорне расподеле одредити *интервале прекривања*, односно интервале у којима се непознати параметар налази са вероватноћом β. Ти интервали су облика (θ_1, θ_2) где су θ_1 и θ_2 одређени из услова $P\{\theta_1 < \theta < \theta_2|x\} = \beta$. Јасно је да овај услов не одређује једнозначно константе θ_1 и θ_2. Често се одређују тако да је $P\{\theta_1 > \theta\} = P\{\theta_2 < \theta\} = \frac{1-\beta}{2}$.