Reinforcement Learning

An Introductory Note

Jingye Wang

wangjy5@shanghaitech.edu.cn

Spring 2020

Contents

2 Review of Basic Probability

For more details of this part one can refer to *Introduction to Probability* [2] and *Monte Carlo Statistical Methods* [3].

2.1 Interpretation of Probability

The Frequentist view: *Probability represents a long-run frequency over a large number of repetitions of an experiment.*

The Bayesian view: *Probability represents a degree of belief about the event in question.*

Many machine learning techniques are derived from these two views. As the computing power and algorithms develop, however, Bayesian is becoming dominant.

2.2 Transformations

Change of variables: *Let* $\boldsymbol{X} = (X_1, ..., X_n)$ *be a continuous random vector with joint PDF* $\boldsymbol{X}$*, and let* $\boldsymbol{Y} = g(\boldsymbol{X})$ *where* g *is an invertible function from* $\mathbb{R}^n$ *to* $\mathbb{R}^n$*. Then the joint PDF of* $\boldsymbol{Y}$ *is*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x}) \left| \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} \right|$$

where the vertical bars say "take the absolute value of the determinant of $\partial\boldsymbol{x}/\partial\boldsymbol{y}$*" and* $\partial\boldsymbol{x}/\partial\boldsymbol{y}$ *is a* **Jacobian matrix**

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}.$$

It assumes that the determinant of the Jacobian matrix is never 0. It also supposes all the partial derivatives $\frac{\partial x_i}{\partial y_j}$ exist and are continuous.

2.3 Limit Theorem

Strong Law of Large Numbers (SLLN): *The sample mean* $\bar{X}_n$ *converges to the true mean* μ *point-wise as* $n \to \infty$*, w.p.1(i.e. with probability 1). In other words, the event* $\bar{X}_n \to \mu$ *has probability 1.*

Weak Law of Large Numbers (WLLN): *For all* $\varepsilon > 0, P(|\bar{X}_n - \mu| > \varepsilon) \to 0$ *as* $n \to \infty$*. (This form of convergence is called* **convergence in probability** *).*

The Weak Law of Large Numbers can be proved by using **Chebyshev's inequality**.

Central Limit Theorem (CLT): *As* $n \to \infty$*,* $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to \mathcal{N}(0, 1)$ *in distribution. In words, the CDF of the left-hand side approaches the CDF of the standard Normal distribution.*

2.4 Sampling & Monte Carlo Methods

Inverse Transform Method: *Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from $(0,1)$ to* R. *We then have the following results.*

1. *Let $U \sim \text{Unif}(0,1)$ and $X = F^{-1}(U)$. Then X is an r.v. with CDF F.*
2. *Let X be an r.v. with CDF F. Then $F(X) \sim \text{Unif}(0,1)$.*

Proof:

1. Let $U \sim \text{Unif}(0,1)$ and $X = F^{-1}(U)$. Then we have $P(U \leq u) = u$ for $u \in (0,1)$. For all real x,

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x),$$

so the CDF of X is F, as claimed.

2. Let X have CDF F, and find the CDF of $Y = F(X)$. Since Y takes values in $(0,1)$, $P(Y \leq y)$ equals 0 for $y \leq 0$ and equals 1 for $y \geq 1$. For $y \in (0,1)$,

$$P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y.$$

Thus Y has the CDF of Unif$(0,1)$. □

Box-Muller: *Let $U \sim \text{Unif}(0,2\pi)$, and let $T \sim \text{Expo}(1)$ be independent of U. Define $X = \sqrt{2T}\cos U$ and $Y = \sqrt{2T}\sin U$. Then X and Y are independent and the joint PDF of $(X,Y) is$*

$$f_{X,Y}(x,y) = \frac{1}{2\pi}e^{\frac{1}{2}(x^2+y^2)}$$

Proof:

The joint PDF of U and T is

$$f_{U,T}(u,t) = \frac{1}{2\pi}e^{-t},$$

for $u \in (0,2\pi)$ and $t > 0$. And we have the Jacobian matrix

$$\frac{\partial(x,y)}{\partial(u,t)} = \begin{pmatrix} -\sqrt{2t}\sin u & \frac{1}{\sqrt{2t}}\cos u \\ \sqrt{2t}\cos u & \frac{1}{\sqrt{2t}}\sin u \end{pmatrix}.$$

Then we have

$$\begin{aligned} f_{X,Y}(x,y) &= f_{U,T}(u,t) \cdot \left| \frac{\partial(u,t)}{\partial(x,y)} \right| \\ &= \frac{1}{2\pi}e^{-t} \cdot 1 \\ &= \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)} \\ &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}}e^{-y^2/2} \end{aligned}$$

for all real x and y. The joint PDF $f_{X,Y}$ factors into a function x times a function of y, so X and Y are independent. Furthermore, we can find that X and Y are i.i.d. $\mathcal{N}(0,1)$. That shows how the *Box-Muller* method works for generating Normal r.v.s. □

Monte Carlo Integration: *Given a function $\Phi : \mathbb{R}^n \to \mathbb{R}$. If $p(\cdot)$ denotes a valid PDF with the support over $\mathbb{R}^n$, then we have*

$$\begin{aligned}\int_{\mathbb{R}^n} \Phi(x)dx &= \int_{\mathbb{R}^n} \frac{\Phi(x)}{p(x)} \cdot p(x)dx \\ &= \mathbb{E}_p\left[\frac{\Phi(x)}{p(x)}\right] \\ &\approx \frac{1}{N}\sum_{k=1}^{N} \frac{\Phi(x_k)}{p(x_k)},\end{aligned}$$

where $x_k \sim p$.

By the law of large numbers, the estimator converges to the true value of the integral with probability 1 as $n \to \infty$. Therefore we can use random samples to obtain approximations of definite integrals when exact integration methods are unavailable. Such approach is often referred to as the **Monte Carlo method**.

Importance Sampling: *Let $p(x)$ denote the target distribution and $\mathbb{E}_p[c(x)]$ is what we want to estimate. With a PDF $q(x)$ which subject to that $\frac{p(x)}{q(x)}$ is finite for all $x \in A$, we have*

$$\begin{aligned}\mathbb{E}_p[c(x)] &= \int_A c(x)p(x)dx \\ &= \int_A c(x) \cdot \frac{p(x)}{q(x)} \cdot q(x)dx \\ &= \mathbb{E}_q\left[c(x) \cdot \frac{p(x)}{q(x)}\right] \\ &\approx \frac{1}{N}\sum_{k=1}^{N} c(x_k)\frac{p(x_k)}{q(x_k)}\end{aligned}$$

where $x_k \sim q$ and q is called **importance distribution**.

The estimator converges to the true value for the same reason the Monte Carlo method converges. Furthermore, the estimator with importance sampling has the less variance than that of the standard Monte Carlo method.

Acceptance-Rejection Method: *Let $X \sim p$ and $Y \sim q$ from which we can relatively easily generate samples. Then for a constant c such that $c \geq \sup_\zeta \frac{p(\zeta)}{q(\zeta)}$, we can simulate $X \sim p$ with three steps:*

Step 1: Generate $y \sim q$.

Step 2: Generate $u \sim \mathrm{U}(0, 1)$.

Step 3: If $u \leq \frac{p(y)}{cq(y)}$, set $x = y$, otherwise go back to Step 1.

Proof:

We now show how it works.

$$\begin{aligned}
P(X \leq \zeta) &= P\left(Y \leq \zeta | U \leq \frac{p(y)}{cq(y)}\right) \\
&= \frac{P(Y \leq \zeta, U \leq \frac{p(y)}{cq(y)})}{P(U \leq \frac{p(y)}{cq(y)})} \\
&= \frac{\int_0^{\zeta} \int_0^{\frac{p(y)}{cq(y)}} 1du \cdot q(y)dy}{\int_{-\infty}^{+\infty} \int_0^{\frac{p(y)}{cq(y)}} 1du \cdot q(y)dy} \\
&= \frac{\int_{-\infty}^{\zeta} p(y)dy}{\int_{-\infty}^{+\infty} p(y)dy}.
\end{aligned}$$

Given that p is a valid PDF, the denominator is 1. Thus we have

$$P(X \leq \zeta) = \int_{-\infty}^{\zeta} p(x)dx,$$

which shows that $X \sim p$. □

2.5 Basic Inequalities

Cauchy-Schwarz Inequality: *For any r.v.s X and Y with finite variances,*

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

Proof:

For any t,

$$\begin{aligned}
\mathbb{E}[(Y - tX)^2] &\geq 0 \\
\mathbb{E}[(Y^2 - 2tXY + t^2X^2)] &\geq 0,
\end{aligned}$$

where the left-hand side is a quadratic function with respect to t. To satisfy the inequality, the discriminant of the quadratic must be less than 0, which means

$$\begin{aligned}
[2\mathbb{E}[XY]]^2 - 4 \cdot \mathbb{E}[X^2]\mathbb{E}[Y^2] &\leq 0 \\
[\mathbb{E}[XY]]^2 &\leq \mathbb{E}[X^2]\mathbb{E}[Y^2] \\
|\mathbb{E}[XY]| &\leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.
\end{aligned}$$

Therefore we have the Cauchy-Schwarz inequality. □

Jensen's Inequality: *Let X be a random variable. If g is a convex function, then $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$. If g is a concave function, then $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$. In both cases, the only way that equality can hold is if there are constants a and b such that $g(X) = a + bX$ with probability 1.*

Proof:

If g is convex, then all lines that are tangent to g lie below g. Denoting the tangent line of g by $a + bx$, we have $g(x) \geq a + bx$ for all x by convexity, so $g(X) \geq a + bX$. Taking the expectation of both sides,

$$\begin{aligned}
\mathbb{E}[g(X)] &\geq \mathbb{E}[a + bX] \\
&\geq a + b\mathbb{E}[X] \\
&\geq g(\mathbb{E}[X]).
\end{aligned}$$

If g is concave, then $h = -g$ is convex, so we can apply the proof to h to see that the inequality for g is reversed from the convex case.

Lastly, assume that $g(X) = a + bX$ holds in the convex case. Let $Y = g(X) - a - bX$. Then Y must be a nonnegative r.v. with $\mathbb{E}[Y] = 0$, so $P(Y = 0) = 1$. So equality holds if and only if $P(g(X) = a+bX) = 1$. For the concave case, the similar argument can apply to $Y = a+bX-g(X)$. $\square$

Norm Inequality: *For a random variable X whose moment of order $r > 0$ is finite, we define the following norm*

$$\|X\|_r = (\mathbb{E}[|X|^r])^{\frac{1}{r}}.$$

With this definition, we have the following inequalities.

- *Holder Inequality*: *Let $\frac{1}{p} + \frac{1}{q} = 1$. If $\mathbb{E}[|X|^p], \mathbb{E}[|X|^q] < \infty$, then $|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq \|X\|_p \cdot \|Y\|_q$.*
- *Lyapunov Inequality*: *For $0 < r \leq p$, $\|X\|_r \leq \|X\|_p$.*
- *Minkowski Inequality*: *Let $p \geq 1$. If $\mathbb{E}[|X|^p], \mathbb{E}[|Y|^p] < \infty$, then $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$.*

Markov's Inequality: *For any r.v. X and constant $a > 0$,*

$$P(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}.$$

Proof:

Let $Y = \frac{|X|}{a}$. We need to show that $P(Y \geq 1) \leq \mathbb{E}[Y]$. Note that

$$I(Y \geq 1) \leq Y,$$

taking the expectation of both sides, then we have Markov's Inequality. $\square$

Markov's inequality is a very crude bound because it requires absolutely no assumptions about X. The right-hand side of the inequality could be greater than 1 sometimes, or even infinite.

Chebyshev's Inequality: *Let X have mean μ and variance σ^2. Then for any $a > 0$,*

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Proof:

By Markov's inequality,

$$\begin{aligned} P(|X-\mu| \geq a) &= P((X-\mu)^2 \geq a^2) \\ &\leq \frac{\mathbb{E}[(X-\mu)^2]}{a^2} \\ &= \frac{\sigma^2}{a^2}. \end{aligned}$$

□

Chernoff's Inequality: *For any r.v. X and constants $a > 0$, $t > 0$, we have*

$$P(X \geq a) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}.$$

Proof:

The transformation g with $g(x) = e^{tx}$ is invertible and strictly increasing. So by Markov's inequality, we have

$$\begin{aligned} P(X \geq a) &= P(e^{tX} \geq e^{ta}) \\ &\leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}. \end{aligned}$$

□

2.6 Concentration Inequalities

Hoeffding Lemma: *Let X be a r.v. with $\mathbb{E}[X] = 0$, taking values in a bounded interval $[a, b]$, where a and b are constants. Then for any $\lambda > 0$,*

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{1}{8}\lambda^2(b-a)^2}.$$

Proof:

For the case $a = b = 0$, we have $P(X = 0) = 1$. The equality holds since both sides of the inequality are 1. We now consider the general case where $a < 0$ and $b > 0$.

Let $f(x) = e^{\lambda x}$ where $x \in [a, b]$. According to its convexity, for any $\alpha \in (0, 1)$, we have

$$f(\alpha a + (1-\alpha)b) \leq \alpha f(a) + (1-\alpha)f(b) = \alpha e^{\lambda a} + (1-\alpha)e^{\lambda b}.$$

As $X \in [a, b]$, let $\alpha = \frac{b-X}{b-a}$, then we have $f(\alpha a + (1-\alpha)b) = f(X) = e^{\lambda X}$. Plugging the two equations into the previous inequality, we have

$$e^{\lambda X} \leq \frac{b-X}{b-a}e^{\lambda a} + \frac{X-a}{b-a}e^{\lambda b}.$$

Taking the expectation of both sides,

$$\mathbb{E}[e^{\lambda X}] \leq \frac{b}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} = e^{\lambda a}\left[\frac{b}{b-a} - \frac{a}{b-a}e^{\lambda(b-a)}\right],$$

and defining a function $\Phi(t) = -\theta t + \ln(1 - \theta + \theta e^t)$ where $\theta = \frac{-a}{b-a} > 0$, we have

$$\mathbb{E}[e^{\lambda X}] \leq e^{\Phi(\lambda(b-a))},$$

as $e^{\Phi(\lambda(b-a))} = e^{\lambda a}\left[\frac{b}{b-a} - \frac{a}{b-a}e^{\lambda(b-a)}\right]$.

We now focus on $\Phi(t)$. According to Taylor expansion, for any $t > 0$, $\exists \tau \in [0, t)$ s.t.

$$\begin{aligned}\Phi(t) &= \Phi(0) + t\Phi'(0) + \frac{1}{2}t^2\Phi''(\tau) \\ &= \frac{1}{2}t^2 \cdot \frac{(1-\theta)\theta e^\tau}{(1-\theta+\theta e^\tau)^2} \\ &\leq \frac{1}{8}t^2\end{aligned}$$

since

$$(1-\theta+\theta e^\tau)^2 = (1-\theta-\theta e^\tau)^2 + 4(1-\theta)\theta e^\tau \geq 4(1-\theta)\theta e^\tau.$$

Plugging in $t = \lambda(b-a)$, we have $\Phi(\lambda(b-a)) \leq \frac{1}{8}\lambda^2(b-a)^2$. It follows that

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{1}{8}\lambda^2(b-a)^2}.$$

□

Hoeffding Bound: *Let $X_1, X_2, ..., X_n$ be independent r.v.s with $\mathbb{E}[X_i] = \mu, a \leq X_i \leq b$ for each $i = 1, 2, ..., n$, where a, b are constants. Then for any $\varepsilon \geq 0$,*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \leq 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}.$$

Proof:

Let $Z_i = X_i - \mu$ and $Z = \frac{1}{n}\sum_{i=1}^n Z_i$, then we have $\mathbb{E}[Z_i] = 0$ and $\mathbb{E}[Z] = 0$. For any $\lambda > 0$, we have

$$P(Z \geq \varepsilon) \leq \frac{\mathbb{E}[e^{\lambda Z}]}{e^{\lambda\varepsilon}}$$

by Chernoff's inequality. For $\mathbb{E}[e^{\lambda Z}]$, we have

$$\mathbb{E}[e^{\lambda Z}] = \mathbb{E}[e^{\lambda \frac{1}{n}\sum_{i=1}^n Z_i}] = \prod_{i=1}^n \mathbb{E}[e^{\frac{\lambda}{n}Z_i}].$$

As $Z_i = X_i - \mu \in [a-\mu, b-\mu]$, using Hoeffding Lemma, we have

$$\prod_{i=1}^n \mathbb{E}[e^{\frac{\lambda}{n}Z_i}] \leq e^{\frac{\lambda^2}{8n}(b-a)^2}.$$

Therefore we have

$$P(Z \geq \varepsilon) \leq e^{-\lambda\varepsilon + \frac{\lambda^2}{8n}(b-a)^2}.$$

Now we focus on the quadratic $-\lambda\varepsilon + \frac{\lambda^2}{8n}(b-a)^2$ w.r.t λ. It is easy to find the quadratic has the minimum at $\lambda = \frac{4n\varepsilon}{(b-a)^2}$. Plugging in $\lambda = \frac{4n\varepsilon}{(b-a)^2}$, we have

$$P(Z \geq \varepsilon) \leq e^{\frac{-2n\varepsilon^2}{(b-a)^2}}.$$

Therefore by the symmetry we have

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \leq 2e^{\frac{-2n\varepsilon^2}{(b-a)^2}}.$$

□

2.7 Conditional Expectation

"Conditional probabilities are probabilities, and all probabilities are conditional."

For conditional expectation, the case is similar.

Taking out what's known: *For any function h,*

$$\mathbb{E}[h(X)Y|X] = h(X)\mathbb{E}[Y|X].$$

Intuitively, when we take expectations given X, we are treating X as if it has crystallized into a known constant. Then any function of X, say h(X), also acts like a known constant while we are conditioning on X.

Law of Total Expectation (LOTE): *Let $A_1, ..., A_n$ be a partition of a sample space, with $P(A_i) > 0$ for all i, and let Y be a random variable on this sample space. Then*

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[Y|A_i]P(A_i).$$

Adam's Law: *For any r.v.s X and Y,*

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

Proof:

Without loss of generality, we consider the case where X and Y are both discrete. Let $\mathbb{E}[Y|X] = g(X)$. Expanding the definition of $g(x)$ by applying LOTUS, we have

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X]] &= \mathbb{E}[g(X)] \\
&= \sum_x g(x)P(X = x) \\
&= \sum_x \mathbb{E}[Y|X = x]P(X = x) \qquad (1) \\
&= \sum_x \left(\sum_y yP(Y = y|X = x) \right) P(X = x) \\
&= \sum_y y \sum_x P(Y = y, X = x) \\
&= \sum_y yP(Y = y) \\
&= \mathbb{E}[Y] \qquad (2)
\end{aligned}$$

as desired. Also, as it shown in the proof, with (1) and (2) we can prove the LOTE. □

Adam's law with extra conditioning: *For any r.v.s X, Y, Z we have*

$$\mathbb{E}[\mathbb{E}[Y|X,Z]|Z] = \mathbb{E}[Y|Z].$$

Proof:

Define the expectation $\hat{\mathbb{E}}[\cdot] = \mathbb{E}[\cdot|Z]$. The key is that *"conditional expectation is expectation"*. We have

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X,Z]|Z] &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[Y|X]] \\
&= \hat{\mathbb{E}}[Y] && \text{(by Adam's Law)} \\
&= \mathbb{E}[Y|Z]
\end{aligned}$$

□

Eve's law: *For any r.v.s X and Y,*

$$Var(Y) = \mathbb{E}[Var(Y|X)] + Var(\mathbb{E}[Y|X]).$$

It is also known as the law of the variance or the variance decomposition formula.

Proof:

Let $g(X) = \mathbb{E}[Y|X]$. By Adam's law, we have $\mathbb{E}[g(X)] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$. According to the variance of the expectation that

$$Var(Y|X) = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2,$$

which can be shown by the way we prove Adam's law, we have

$$\begin{aligned}
\mathbb{E}[Var(Y|X)] &= \mathbb{E}[\mathbb{E}[Y^2|X] - (g(X))^2] \\
&= \mathbb{E}[\mathbb{E}[Y^2|X]] - \mathbb{E}[(g(X))^2] \\
&= \mathbb{E}[Y^2] - \mathbb{E}[(g(X))^2], && (1) \\
Var(\mathbb{E}[Y|X]) &= \mathbb{E}[(g(X))^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 \\
&= \mathbb{E}[g(X)^2] - (\mathbb{E}[Y])^2. && (2)
\end{aligned}$$

Then the Eve's law can be shown by $(1) + (2)$. □