Review.

<u>Def.</u> $\mathbb{E}_{p(x)}[f(\mathbb{X})] := \int f(x)p(x)dx.$

$$\mathbb{V}_{p(x)}[\mathbb{X}] := \mathbb{E}_p[(\mathbb{X}-\mu)(\mathbb{X}-\mu)^T]$$

$$= \mathbb{E}_p[\mathbb{X}\mathbb{X}^T] - \mu\mu^T \qquad (\mu = \mathbb{E}_p[\mathbb{X}])$$

$$H[p(x)] := -\mathbb{E}_p[\log p(\mathbb{X})].$$

$$KL[q(x)\|p(x)] := \mathbb{E}_q[\log q(\mathbb{X})] - \mathbb{E}_q[\log p(\mathbb{X})]$$

$$= -H[q(x)] - \mathbb{E}_q[\log p(\mathbb{X})].$$

- サンプリングによる期待値の近似計算.

サンプリング $x^{(1)}, \cdots, x^{(L)} \sim p(x).$

近似 : $\mathbb{E}_p[f(\mathbb{X})] \approx \frac{1}{L}\sum_{l=1}^{L} f(x^{(l)})$

- 大数の法則により, $L \to \infty$ で (r.h.s.) $\xrightarrow{a.s.}$ (l.h.s.)

2.2 離散確率分布.

* Bernoulli分布.

pmf. $\mathrm{Bern}(x|\mu) = \mu^x (1-\mu)^{1-x}$.

・平均.

$$\mathbb{E}[X] = \sum_{x=0}^{1} x \mu^x (1-\mu)^{1-x}$$
$$= \mu^1 (1-\mu)^{1-1} = \mu.$$

・分散.

$$\mathbb{V}[X] = \sum_{x=0}^{1} x^2 \mu^x (1-\mu)^{1-x} - \mathbb{E}[X]^2$$
$$= \mu - \mu^2 = \mu(1-\mu).$$

・エントロピー

$$H[\mathrm{Bern}(x|\mu)] = -\mathbb{E}[\log \mathrm{Bern}(X|\mu)]$$
$$= -\mathbb{E}[X \log \mu + (1-X)\log(1-\mu)]$$
$$= -\mathbb{E}[X]\log \mu - \mathbb{E}[1-X]\log(1-\mu)$$
$$= -\mu \log \mu - (1-\mu)\log(1-\mu).$$

・$\mu = 0.5$ でエントロピー最大.

- KL divergence.

$$p(x) = \mathrm{Bern}(x|\mu), \quad q(x) = \mathrm{Bern}(x|\hat{\mu}).$$

$$\mathrm{KL}[q(x) \| p(x)]$$

$$= -H[q(x)] - \mathbb{E}_q[\log p(X)]$$

$$\mathbb{E}_q[\log p(X)] = \mathbb{E}_q[X \log\mu + (1-X)\log(1-\mu)]$$

$$= \mathbb{E}_q[X]\log\mu + \mathbb{E}_q[1-X]\log(1-\mu)$$

$$= \hat{\mu}\log\mu + (1-\hat{\mu})\log(1-\mu).$$

$$\therefore \mathrm{KL}[q(x)\|p(x)] = \hat{\mu}\log\frac{\hat{\mu}}{\mu} + (1-\hat{\mu})\log\frac{1-\hat{\mu}}{1-\mu}.$$

* 二項分布.

pmf $\quad \mathrm{Bin}(m|M,\mu) = \binom{M}{m}\mu^m(1-\mu)^{M-m}$

- 平均.

$$\mathbb{E}[X] = \sum_{m=0}^{M} m\binom{M}{m}\mu^m(1-\mu)^{M-m}$$

$$= \sum_{m=1}^{M} m\,\frac{M!}{m!\,(M-m)!}\mu^m(1-\mu)^{M-m}$$

$$= \sum_{m=1}^{M}\frac{M\cdot(M-1)!}{(m-1)!\,(M-m)!}\mu\cdot\mu^{m-1}(1-\mu)^{M-m}$$

$$= M\mu \sum_{m=1}^{M-1} \frac{(M-1)!}{(m-1)!(M-m)!} \mu^{m-1} (1-\mu)^{M-m}$$

$m' := m-1$ とおく.
$m = m'+1$.

$$= M\mu \sum_{m'=0}^{M-1} \frac{(M-1)!}{m'!(M-1-m')!} \mu^{m'} (1-\mu)^{M-1-m'}$$

$$= M\mu \underbrace{\sum_{m'=0}^{M-1} \mathrm{Bin}(m' \mid M-1, \mu)}_{=1}$$

$$= M\mu.$$

・ 分散.

$\mathbb{V}[X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2$ を使う方が楽かも.

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

$$\mathbb{E}[X^2] = \sum_{m=0}^{M} m^2 \binom{M}{m} \mu^m (1-\mu)^{M-m}$$

$$= M\mu \sum_{m=1}^{M} m \frac{(M-1)!}{(m-1)!(M-m)!} \mu^{m-1} (1-\mu)^{M-m}$$

$$= M\mu \sum_{m'=0}^{M-1} (m'+1) \frac{(M-1)!}{m'!(M-1-m')!} \mu^{m'} (1-\mu)^{M-1-m'}$$

$$= M\mu \left(\sum_{m'=0}^{M-1} m' \mathrm{Bin}(m' \mid M-1, \mu) + \underbrace{\sum_{m'=0}^{M-1} \mathrm{Bin}(m' \mid M-1, \mu)}_{=1} \right)$$

$$= M\mu \left(\mathbb{E}_{\mathrm{Bin}(m' \mid M-1, \mu)}[X] + 1 \right)$$

$$= M\mu \left((M-1)\mu + 1 \right).$$

$$\therefore \mathbb{V}[X] = M\mu(1-\mu).$$

* カテゴリー分布.

$$\mathbf{s} \in \{0,1\}^K. \quad \sum_{k=1}^{K} s_k = 1 \quad (\mathbf{s} \text{ は1つだけ1のベクトル})$$

pmf. $$\mathrm{Cat}(\mathbf{s} \mid \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{s_k}$$

$$\left(\text{param} : \boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_K \end{pmatrix}, \quad \boldsymbol{\pi} \text{ は prob.vec.} \right)$$

・平均

$$\mathcal{S} = \{ \mathbf{e}_1, \cdots, \mathbf{e}_K \} \qquad \mathbf{e}_i = (\delta_{ij})_{j\downarrow}$$

$$\mathbb{E}[\mathbf{S}] = \sum_{\mathbf{s}\in\mathcal{S}} \mathbf{s} \prod_{k=1}^{K} \pi_k^{s_k}$$

$$= \sum_{k=1}^{K} \pi_k \mathbf{e}_k = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_K \end{pmatrix} = \boldsymbol{\pi}.$$

・分散.

$$\mathbb{V}[\mathbf{S}] = \mathbb{E}[\mathbf{S}\mathbf{S}^T] - \mathbb{E}[\mathbf{S}]\mathbb{E}[\mathbf{S}]^T$$

$$\mathbb{E}[\mathbf{S}\mathbf{S}^T] = \sum_{\mathbf{s}\in\mathcal{S}} \mathbf{s}\mathbf{s}^T \prod_{k=1}^{K} \pi_k^{s_k}$$

$$= \sum_{k=1}^{K} \mathbf{e}_k \mathbf{e}_k^T \pi_k = \mathrm{diag}(\boldsymbol{\pi}).$$

$$\therefore \mathbb{V}[\mathbf{s}] = \mathrm{diag}(\pi_1, \cdots, \pi_K) - \boldsymbol{\pi}\boldsymbol{\pi}^\top .$$

$$= \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_K \\ & \ddots & \ddots & \vdots \\ \mathrm{Sym.} & & \ddots & -\pi_{K-1}\pi_K \\ & & & \pi_K(1-\pi_K) \end{pmatrix}$$

・エントロピー

$$H[\mathrm{Cat}(\mathbf{s}|\boldsymbol{\pi})] = -\mathbb{E}[\log \mathrm{Cat}(\mathbf{s}|\boldsymbol{\pi})]$$

$$= -\mathbb{E}\left[\sum_{k=1}^{K} s_k \log \pi_k\right]$$

$$= -\mathbb{E}[\mathbf{s}^\top \log \boldsymbol{\pi}] \qquad \log \boldsymbol{\pi} = \begin{pmatrix} \log \pi_1 \\ \vdots \\ \log \pi_K \end{pmatrix}$$

$$= -\boldsymbol{\pi}^\top \log \boldsymbol{\pi} .$$

＊ 多項分布.

pmf. $$\mathrm{Mult}(\mathbf{m}|\boldsymbol{\pi}, M) = M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$\mathbf{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_K \end{pmatrix}, \quad \sum_{k=1}^{K} m_k = M, \quad m_k \in \{0, \cdots, M\}$$

・平均.

$$\mathcal{M}_M := \left\{ \mathbf{m} \;\middle|\; \sum_{k=1}^{K} m_k = M, \ m_k \in \{0, \cdots, M\} \right\}$$ として,

$$\mathbb{E}[\mathbf{x}] = \sum_{\mathbf{m} \in \mathcal{M}_M} \mathbf{m}\, M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$= \begin{pmatrix} \displaystyle\sum_{m_1=1}^{M} m_1 \frac{M!}{m_1!(M-m_1)!} \pi_1^{m_1} (1-\pi_1)^{M-m_1} \\ \vdots \\ \displaystyle\sum_{m_K=1}^{M} m_K \frac{M!}{m_K!(M-m_K)!} \pi_1^{m_K} (1-\pi_1)^{M-m_K} \end{pmatrix}$$

$$= M\boldsymbol{\pi}.$$

・分散.

$$\mathbb{V}[\mathbf{x}] = \mathbb{E}[\mathbf{x}\mathbf{x}^T] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]^T.$$

$$\mathbb{E}[\mathbf{x}\mathbf{x}^T] = \sum_{\mathbf{m} \in \mathcal{M}_M} \mathbf{m}\mathbf{m}^T M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$= \sum_{\mathbf{m} \in \mathcal{M}_M} \begin{pmatrix} m_1^2 & m_1 m_2 & \cdots & m_1 m_K \\ & \ddots & \ddots & \vdots \\ \mathrm{Sym.} & & \ddots & m_{K-1} m_K \\ & & & m_K^2 \end{pmatrix} M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

対角要素.

$$\sum_{\mathbf{m} \in \mathcal{M}_M} m_k^2 M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$= \sum_{m_k=0}^{M} m_k^2 \frac{M!}{m_k!(M-m_k)!} \pi_k^{m_k} (1-\pi_k)^{M-m_k}$$

$$= M\pi_k\left((M-1)\pi_k + 1\right)$$

2項分布の分散のときに計算している.

非対角要素. $(j \neq k.)$

$$\sum_{m \in \mathcal{M}_M} m_j m_k M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$= \sum_{m_j=0}^{M} \sum_{m_k=0}^{M-m_j} m_j m_k M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$= \sum_{m_j=0}^{M} \sum_{m_k=0}^{M-m_j} m_j m_k \frac{M!}{m_j!\, m_k!\, (M-m_j-m_k)!} \pi_j^{m_j} \pi_k^{m_k} (1-\pi_j-\pi_k)^{M-m_j-m_k}$$

$m_j + m_k =: r$ とおく

$$= \sum_{r=0}^{M} \sum_{m_j=0}^{r} m_j (r-m_j) \frac{M!}{r!(M-r)!} \frac{r!}{m_j!(r-m_j)!} \pi_j^{m_j} \pi_k^{r-m_j} (1-\pi_j-\pi_k)^{M-}$$

$$= \pi_j \pi_k \sum_{r=0}^{M} \frac{M!\, r(r-1)}{r!(M-r)!} (1-\pi_j-\pi_k)^{M-r} \underline{\sum_{m_j=1}^{r-1} \binom{r-2}{m_j-1} \pi_j^{m_j-1} \pi_k^{r-m_j-1}}$$

2項定理.

$$= \pi_j \pi_k \sum_{r=0}^{M} \frac{M!\, r(r-1)}{r!(M-r)!} (1-\pi_j-\pi_k)^{M-r} \underline{(\pi_j+\pi_k)^{r-2}}$$

$$= \pi_j \pi_k \sum_{r=2}^{M} \frac{M!}{(r-2)!(M-r)!} (\pi_j+\pi_k)^{r-2} (1-(\pi_j+\pi_k))^{M-r}$$

$$= \pi_j \pi_k \sum_{r'=0}^{M-2} \frac{M!}{r'!(M-2-r')!} (\pi_j+\pi_k)^{r'} (1-(\underbrace{\pi_j+\pi_k}_{=:p}))^{M-2-r'}$$

$$= \pi_j \pi_k M(M-1) \underbrace{\sum_{r'=0}^{M-2} \binom{M-2}{r'} p^{r'} (1-p)^{M-2-r'}}_{=1}$$

$$= M(M-1)\pi_j \pi_k.$$

Remark 次のようにも計算できる(し、こちらの方が楽かも)

$$\sum_{\mathbf{m}\in\mathcal{M}_M} m_j m_k M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!}$$

$$= \sum_{\substack{\mathbf{m}\in\mathcal{M}_M,\\ m_j\neq 0,\, m_k\neq 0}} \frac{M(M-1)\cdot(M-2)!}{m_1!\cdots(m_j-1)!\cdots(m_k-1)!\cdots m_K!} \pi_1^{m_1}\cdot\cdots\cdot\pi_j^{m_j-1+1}\cdot\cdots\cdot\pi_k^{m_k-1+1}\cdot\cdots\cdot\pi_K^{m_K}$$

$$= M(M-1)\pi_j\pi_k \sum_{\mathbf{m}'\in\mathcal{M}_{M-2}} (M-2)! \prod_{k=1}^{K} \frac{\pi_k^{m_k'}}{m_k'!}$$

$$= M(M-1)\pi_j\pi_k \underbrace{\sum_{\mathbf{m}'\in\mathcal{M}_{M-2}} \mathrm{Mult}(\mathbf{m}'\mid\boldsymbol{\pi}, M-2)}_{=1} = M(M-1)\pi_j\pi_k$$

$$\mathbb{E}[\mathbb{X}]\mathbb{E}[\mathbb{X}]^T = M^2\boldsymbol{\pi}\boldsymbol{\pi}^T = M^2\begin{pmatrix} \pi_1^2 & \pi_1\pi_2 & \cdots & \pi_1\pi_K \\ & \ddots & \ddots & \vdots \\ & & \ddots & \pi_{K-1}\pi_K \\ \mathrm{Sym.} & & & \pi_K^2 \end{pmatrix}$$

$\therefore \mathbb{V}[X]$ の対角要素:

$$M\pi_k((M-1)\pi_k+1) - M^2\pi_k^2 = M\pi_k(1-\pi_k)$$

$\mathbb{V}[X]$ の非対角要素:

$$M(M-1)\pi_j\pi_k - M^2\pi_j\pi_k = -M\pi_j\pi_k.$$

$$\therefore \mathbb{V}[\mathbb{X}] = \begin{pmatrix} M\pi_1(1-\pi_1) & -M\pi_1\pi_2 & \cdots & -M\pi_1\pi_K \\ & \ddots & \ddots & \vdots \\ & & \ddots & -M\pi_{K-1}\pi_K \\ \mathrm{Sym.} & & & M\pi_K(1-\pi_K) \end{pmatrix}$$

$$= M(\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T).$$

* Poisson分布.

pmf. $\mathrm{Poi}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (\lambda > 0)$.

・ 対数表示:

$$\log \mathrm{Poi}(x|\lambda) = x\log\lambda - \log x! - \lambda.$$

・ 平均.

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda}$$

$$= \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} e^{-\lambda}$$

$x' := x-1$ とおく.

$$= \sum_{x'=0}^{\infty} \frac{\lambda^{x'+1}}{x'!} e^{-\lambda}$$

$$= \lambda \underbrace{\sum_{x'=0}^{\infty} \mathrm{Poi}(x'|\lambda)}_{=1} = \lambda.$$

・ 分散.

$\mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2$ の方が楽.

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

$$\mathbb{E}[X^2] = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda}$$

$$= \sum_{x=1}^{\infty} x \frac{\lambda^x}{(x-1)!} e^{-\lambda}$$

$$= \sum_{x'=0}^{\infty} (x'+1) \frac{\lambda^{x'+1}}{x'!} e^{-\lambda}$$

$$= \lambda \sum_{x'=0}^{\infty} x' \mathrm{Poi}(x'|\lambda) + \lambda \sum_{x'=0}^{\infty} \mathrm{Poi}(x'|\lambda)$$

$$= \lambda \mathbb{E}[X] + \lambda = \lambda^2 + \lambda .$$

$$\therefore \mathbb{V}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda .$$