Bernoulli distribution

The Bernoulli distribution for a single binary random variable X with state $x \epsilon \{0,1\}$.

It is governed by a single continuous parameter $\mu \epsilon [0,1]$ that represents the probability of $X = 1$.

The distribution $Ber(\mu) is\ defined\ as$:

$$P(x|\mu) = \mu^x(1-\mu)^{1-x}, x\ \epsilon\{0,1\},$$

$$E[x] = \ \mu,$$

$$V[x] = \mu(1-x),$$

Where $E[x]$ and $V[x]$ are the mean and variance of binary random variable.

Exponential Family

$$\boldsymbol{f(x|\theta) = h(x)exp\big(\eta(\theta)T(x) - A(\theta)\big)}$$

Now , take the Bernoulli distribution series:

$P(x|\mu) = \mu^x(1-\mu)^{1-x}$

Note:

$$\boldsymbol{f(x|\mu) = \begin{cases} \mu, & if\ x\ =\ 1 \\ 1\ -\ \mu, & if\ x\ =\ 0 \end{cases}}$$

$\boldsymbol{i.e.\,\mu\ when\ x = 1\ i.e.\,success\ and\ 1 - \ \mu, if\ x = 0\ i.e.\,failures\ and\ P(x|\mu) where\ P\ stands\ for\ probability}.$

Now let us expand the Bernoulli's distribution,

$$P(x|\mu) = \mu^x(1-\mu)^{1-x}$$

$$=> exp\left[log\left(\mu^x(1-\mu)^{1-x}\right)\right]$$

$$=> exp[xlog(\mu) + (1-x)log(1-\mu)]$$

$$=> exp[xlog(\mu) + log(1-\mu) - xlog(1-\mu)]$$

$$=> exp[xlog(\mu) - xlog(1-\mu) + log(1-\mu)]$$

$$=> exp\left[xlog(\frac{\mu}{1-\mu}) + log(1-\mu)\right] - eqn(i)$$

Note , the above eqn(i) can be interpreted in different ways:

i.e. if $\boldsymbol{\mu = P, then}$:

$$exp\left[xlog(\frac{p}{1-p}) + log(1-p)\right]$$

And during that time Bernoulli's distribution becomes:

$$P(x|p) = p^x(1-p)^{1-x}$$

And,

$$\boldsymbol{f(x|p)} = \begin{cases} \boldsymbol{p}, & \boldsymbol{if\ x\ =\ 1} \\ \boldsymbol{1\ -\ p}, & \boldsymbol{if\ x\ =\ 0} \end{cases}$$

Therefore, it depends on notable variable we use in the equation.

Now we can write the above equation as:

$$\boldsymbol{exp\left[xlog\left(\frac{\mu}{1-\mu}\right) - (-log(1-\mu))\right] - eqn(ii)}$$

Now,

$$h(x) = 1,$$

$$\theta = \log\left(\frac{\boldsymbol{\mu}}{\boldsymbol{1-\mu}}\right),$$

$$\theta(x) = 1 \times x = x,$$

$$A(\theta) = -\log(1-\mu)$$

The above interpretation is based on exponential family,

$$\boldsymbol{f(x|\theta) = h(x)exp\big(\eta(\theta)T(x) - A(\theta)\big)}$$

i.e.

$$h(x) = 1,$$

$$\eta(\theta) = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right),$$

$$T(x) = 1 \times x = x,$$

$$A(\theta) = -\log(1 - \mu)$$

Now, note the relationship between $\mu \ and \ \theta$ $are \ invertible$:

$$\mu = \frac{1}{1 + \exp^{-\theta}} - eqn(i)$$

And,

$$\exp^{-\theta} = \frac{1 - \mu}{\mu} - eqn(a)$$

Or,

$$\mu = \frac{1}{1 + \frac{1}{\exp^{\theta}}}$$

$$=> \mu = \frac{1}{\frac{\exp^{\theta} + 1}{\exp^{\theta}}}$$

$$=> \mu = \frac{\exp^{\theta}}{\exp^{\theta} + 1} - eqn(ii)$$

Similarly:

$$\exp^{\theta} = \frac{\mu}{1-\mu} - eqn(b)$$

From the above equation(i) we get :

$$\mu = \frac{1}{1+\exp^{-\theta}}$$

$$=> \mu\left(1+\exp^{-\theta}\right) = 1$$

$$=> \left(1+\exp^{-\theta}\right) = \frac{1}{\mu}$$

$$=> \left(\exp^{-\theta}\right) = \frac{1}{\mu} - 1$$

$$=> \left(\exp^{-\theta}\right) = \frac{1-\mu}{\mu}$$

$$=> \mu\left(\exp^{-\theta}\right) = 1-\mu$$

we know ,

$$\mu = \frac{1}{1+\exp^{-\theta}}$$

Therefore,

$$=> \frac{1}{1+\exp^{-\theta}} \times \left(\exp^{-\theta}\right) = 1-\mu$$

$$=> \frac{\exp^{-\theta}}{1+\exp^{-\theta}} = 1-\mu$$

we get,

$$=> \frac{\frac{1}{\exp^{\theta}}}{1 + \frac{1}{\exp^{\theta}}} = 1 - \mu$$

$$=> \frac{\frac{1}{\exp^{\theta}}}{\frac{\exp^{\theta} + 1}{\exp^{\theta}}} = 1 - \mu$$

$$=> \frac{1}{\exp^{\theta}} \times \frac{\exp^{\theta}}{\exp^{\theta} + 1} \ = 1 - \mu$$

$$=> \frac{1}{\exp^{\theta} + 1} \ = 1 - \mu$$

Therefore we got,

$$\frac{1}{\exp^{\theta} + 1} \ = 1 - \mu$$

Putting it in:

$$\boldsymbol{exp\left[xlog\left(\frac{\mu}{1-\mu}\right) - (-log(1-\mu))\right]}$$

we get,

$$=> \boldsymbol{exp}\left[\boldsymbol{xlog}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right) - \left(-\left(\boldsymbol{log}\left(\frac{1}{\exp^{\theta}+1}\right)\right)\right)\right]$$

$$=> \boldsymbol{exp}\left[\boldsymbol{xlog}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right) - \left(-\left(\boldsymbol{log}(\mathbf{1}) - \boldsymbol{log}\left(\exp^{\theta}+1\right)\right)\right)\right]$$

$$=> \boldsymbol{exp}\left[\boldsymbol{xlog}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right) - \left(-\left(\mathbf{0} - \boldsymbol{log}(\exp^{\theta}+1)\right)\right)\right]$$

$$=> \boldsymbol{exp}\left[\boldsymbol{xlog}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right) - \left(-\left(-\boldsymbol{log}(\exp^{\theta}+1)\right)\right)\right]$$

$$=> \boldsymbol{exp}\left[\boldsymbol{xlog}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right) - \left(+\boldsymbol{log}\left(\exp^{\theta}+1\right)\right)\right]$$

In Canonical Form

$$\boldsymbol{f(x|\theta) = h(x)exp\left(\eta(\theta)T(x) - A(\theta)\right)}$$

$$h(x) = 1,$$

$$\eta(\theta) = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right),$$

$$T(x) = x,$$

$$A(\theta) = \boldsymbol{log}\left(\exp^{\theta}+1\right)$$

Mean

$$E\big(T(x)\big) = E[x]$$

$$= \frac{\partial}{\partial\theta} A(\theta)$$

$$= \frac{\partial}{\partial\theta} \times \boldsymbol{log}\big(\exp^{\theta} + 1\big)$$

$$= \frac{1}{1 + \exp^{\theta}} \times \frac{\partial}{\partial\theta}(1 + \exp^{\theta})$$

$$= \frac{1}{1 + \exp^{\theta}} \times (0 + \frac{\partial}{\partial\theta} \times \exp^{\theta})$$

Note, ‘**exp’** represents as ‘**e’** and ‘**exp’** stands for ‘**exponent’**.

Re-writing the equation:

$$= \frac{1}{1 + \mathrm{e}^{\theta}} \times (0 + \frac{\partial}{\partial\theta} \times \mathrm{e}^{\theta})$$

$$= \frac{1}{1 + \mathrm{e}^{\theta}} \times (0 + \mathrm{e}^{\theta})$$

$$= \frac{1}{1 + \mathrm{e}^{\theta}} \times (\mathrm{e}^{\theta})$$

$$= \frac{\mathrm{e}^{\theta}}{1 + \mathrm{e}^{\theta}}$$

We know,

$$\theta = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right),$$

Or,

$$\eta(\theta) = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right)$$

Hence putting the value of θ in the equation:

$$= \frac{\mathrm{e}^{\left(\log\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)\right)}}{1 + \mathrm{e}^{\left(\log\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)\right)}}$$

We can also represent it as:

$$= \frac{\mathrm{e}^{\left(\log_{\mathrm{e}}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)\right)}}{1 + \mathrm{e}^{\left(\log_{\mathrm{e}}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)\right)}}$$

$$we\ know, a^{\log_{\mathrm{a}} b} = b$$

$$= \frac{\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}{1 + \frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}$$

$$= \frac{\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}{\frac{\mathbf{1} - \boldsymbol{\mu} + \boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}$$

$$= \frac{\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}{\frac{\mathbf{1}}{\mathbf{1} - \boldsymbol{\mu}}}$$

$$= \frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}} \times \frac{\mathbf{1} - \boldsymbol{\mu}}{\mathbf{1}}$$

$$= \boldsymbol{\mu}$$

Hence, Mean = $E[x] = \boldsymbol{\mu}$

Variance

$$Var\big(T(x)\big) = Var[x]$$

We know variance = σ^2

$$Var[x] = \frac{\partial^2}{\partial\theta^2} \times A(\theta)$$

$$= \frac{\partial^2}{\partial\theta^2} \times \Big(\boldsymbol{log}\big(\mathrm{exp}^{\theta} + 1\big)\Big)$$

$$= \frac{\partial}{\partial\theta} \times \frac{\partial}{\partial\theta} \times \left(\boldsymbol{log}\left(\exp^{\theta} + 1\right)\right)$$

$\boldsymbol{We\ know}\ \frac{\partial}{\partial\theta} \times \left(\boldsymbol{log}\left(\exp^{\theta} + 1\right)\right) = \frac{\mathrm{e}^{\theta}}{1 + \mathrm{e}^{\theta}}$

$as\ we\ got\ it\ while\ doing\ calculation\ of\ mean, therefore$:

$$= \frac{\partial}{\partial\theta} \times \left(\frac{\mathrm{e}^{\theta}}{1 + \mathrm{e}^{\theta}}\right)$$

We know the division calculation of differential calculus: $\frac{\partial}{\partial x} \times \frac{f(x)}{g(x)} = \frac{g(x) \times \frac{\partial}{\partial x} \times f(x) - f(x) \times \frac{\partial}{\partial x} \times g(x)}{\left(g(x)\right)^2}$

$$= \frac{\left(1 + e^{\theta}\right) \times \frac{\partial}{\partial\theta} \times \mathrm{e}^{\theta} - \left(\mathrm{e}^{\theta} \times \frac{\partial}{\partial\theta} \times \left(1 + e^{\theta}\right)\right)}{(1 + e^{\theta})^2}$$

$$= \frac{\left(1 + e^{\theta}\right) \times \mathrm{e}^{\theta} - \left(\mathrm{e}^{\theta} \times \left(0 + \frac{\partial}{\partial\theta} \times e^{\theta}\right)\right)}{(1 + e^{\theta})^2}$$

$$= \frac{e^{\theta} + e^{2\theta} - \left(\mathrm{e}^{\theta} \times \left(0 + \mathrm{e}^{\theta}\right)\right)}{(1 + e^{\theta})^2}$$

$$= \frac{e^{\theta} + e^{2\theta} - e^{2\theta}}{(1 + e^{\theta})^2}$$

$$= \frac{e^{\theta}}{(1 + e^{\theta})^2}$$

We know,

$$\theta = \log\left(\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right),$$

Therefore,

$$= \frac{e^{\log\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)}}{\left(1 + e^{\log\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)}\right)^2}$$

We can write it as:

$$= \frac{e^{\log_{\mathrm{e}}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)}}{\left(1 + e^{\log_{\mathrm{e}}\left(\frac{\boldsymbol{\mu}}{\mathbf{1}-\boldsymbol{\mu}}\right)}\right)^2}$$

$$we\ know, a^{\log_{\mathrm{a}} b} = b$$

$$= \frac{\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}{\left(1 + \frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right)^2}$$

$$= \frac{\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}{\left(\frac{\mathbf{1} - \boldsymbol{\mu} + \boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}\right)^2}$$

$$= \frac{\frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}}}{\left(\frac{\mathbf{1}}{\mathbf{1} - \boldsymbol{\mu}}\right)^2}$$

$$= \frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}} \times (\mathbf{1} - \boldsymbol{\mu})^{\mathbf{2}}$$

$$= \boldsymbol{\mu} \times (\mathbf{1} - \boldsymbol{\mu})$$

Therefore, **Variance =** $Var[x] = \boldsymbol{\mu} \times (\mathbf{1} - \boldsymbol{\mu})$

Now if we represent $\boldsymbol{\mu}$ $\boldsymbol{as}$ $\boldsymbol{p}$

Then,

Mean = $E[x] = \boldsymbol{p}$

$Var[x] = \boldsymbol{p} \times (\mathbf{1} - \boldsymbol{p})$