Contents

Appendix B
Iterated expectations

Along with Bayes' theorem (the glue holding consistent probability assessment together), iterated expectations is extensively employed for connecting conditional expectation (regression) results with causal effects of interest.

Theorem 11 *Law of iterated expectations*

$$E[Y] = E_X[E[Y \mid X]]$$

Proof.

$$\begin{aligned} E_X[E[Y \mid X]] &= \int_{\underline{x}}^{\overline{x}} E[Y \mid X] f(x)\,dx \\ &= \int_{\underline{x}}^{\overline{x}} \left[\int_{\underline{y}}^{\overline{y}} y f(y \mid x)\,dy \right] f(x)\,dx \end{aligned}$$

By Fubini's theorem, we can change the order of integration

$$\int_{\underline{x}}^{\overline{x}} \left[\int_{\underline{y}}^{\overline{y}} y f(y \mid x)\,dy \right] f(x)\,dx = \int_{\underline{y}}^{\overline{y}} y \left[\int_{\underline{x}}^{\overline{x}} f(y \mid x) f(x)\,dx \right] dy$$

The product rule of Bayes' theorem, $f(y \mid x) f(x) = f(y, x)$, implies iterated expectations can be rewritten as

$$E_X[E[Y \mid X]] = \int_{\underline{y}}^{\overline{y}} y \left[\int_{\underline{x}}^{\overline{x}} f(y, x)\,dx \right] dy$$

Finally, the summation rule integrates out X, $\int_{\underline{x}}^{\overline{x}} f(y, x)\, dx = f(y)$, and produces the result.

$$E_X[E[Y \mid X]] = \int_{\underline{y}}^{\overline{y}} y f(y)\, dy = E[Y]$$

∎

B.1 Decomposition of variance

Corollary 1 *Decomposition of variance.*

$$Var[Y] = E_X[Var[Y \mid X]] + Var_X[E[Y \mid X]]$$

Proof.

$$\begin{aligned}
Var[Y] &= E\left[Y^2\right] - E[Y]^2 \\
&= E_X\left[E\left[Y^2 \mid X\right]\right] - E_X[E[Y \mid X]]^2 \\
Var[Y] &= E_X\left[E\left[Y^2 \mid X\right]\right] - E[Y]^2 \\
&= E_X\left[Var[Y \mid X] + E[Y \mid X]^2\right] - E_X[E[Y \mid X]]^2 \\
&= E_X[Var[Y \mid X]] + E_X\left[E[Y \mid X]^2\right] - E_X[E[Y \mid X]]^2 \\
Var[Y] &= E_X[Var[Y \mid X]] + Var_X[E[Y \mid X]]
\end{aligned}$$

The second line draws from iterated expectations while the fourth line is the decomposition of the second moment. ■

In analysis of variance language, the first term is the residual variation (or variation unexplained)and the second term is the regression variation (or variation explained).

B.2 Jensen's inequality

For any concave function $g(x)$, $E[g(x)] \leq g(E[x])$. Likewise, for any convex function $h(x)$, $E[h(x)] \geq h(E[x])$. Hence, utility functions exhibiting concavity characterize risk aversion or positive risk premia while utility functions exhibiting convexity characterize risk-seeking preferences or negative risk premia.

Further, Jensen's inequality tells us the geometric mean, $G(x)$, is less than or equal to the arithmetic mean, $A(x)$, with equality only when all outcomes are the same.

$$G(x) \equiv \prod_{i=1}^{n} x_i^{p_i} \leq A(x) \equiv \sum_{i=1}^{n} p_i x_i$$

To see this result, let $g(\cdot)$ be the logarithm (a monotone increasing, concave function) for x nonnegative (if any $x_i = 0$ then the inequality is trivially satisfied as the geometric mean is zero if any $x_i = 0$)

$$E[g(x)] = \sum_{i=1}^{n} p_i \ln x_i \leq g(E[x]) = \ln \sum_{i=1}^{n} p_i x_i$$

Let $p_i = \frac{w_i}{w}$ where $w = \sum_{i=1}^{n} w_i$, then

$$\sum_{i=1}^{n} \frac{w_i}{w} \ln x_i \leq \ln \sum_{i=1}^{n} \frac{w_i}{w} x_i$$

To recover geometric and arithmetic means exponentiate (a monotone increasing function) both sides

$$\sqrt[w]{\prod_{i=1}^{n} x_i^{w_i}} = G(x) \leq \sum_{i=1}^{n} \frac{w_i}{w} x_i = A(x)$$